THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :
Systèmes Informatiques Critiques
Systèmes Embarqués

Présentée et soutenue par :
Thierry GERMA

le : vendredi 24 septembre 2010

Titre :

Fusion de données hétérogènes
pour la perception de l'homme
par un robot mobile

Ecole doctorale :
Systèmes (EDSYS)

Unité de recherche :
LAAS-CNRS

Directeur(s) de Thèse :

Frédéric LERASLE

Rapporteurs :
E. Colle (Professeur des Universités, CEMIF)
J.M. Odobez (Senior Researcher, IDIAP)

Autre(s) membre(s) du jury

P. Dalle (Professeur des Universités, IRIT)
M. Devy (Directeur de Recherche, LAAS-CNRS)
W. Puech (Professeur des Universités, LIRMM)
F. Lerasle (Maître de Conférence HDR, LAAS-CNRS)

Remerciements

Je tiens à remercier tout d'abord le LAAS-CNRS, représenté par son directeur Raja Chatilla, et plus particulièrement le groupe RAP, représenté par Michel Devy, pour m'avoir accueilli dans leur structure tout au long de cette aventure professionnelle certe, mais aussi humaine. Tout particulièrement, je tiens à remercier Frédéric Lerasle pour la qualité de son encadrement, son écoute et sa collaboration au sein du projet CommRob. Merci à Patrick Danès et Viviane Cadenat pour leurs conseils avisés, ainsi qu'à Noureddine Ouadah et Adrien Durand-Petiteville pour leur collaboration et leur implication dans mes travaux d'intégration ; une partie des résultats présentés n'auraient sûrement pas vu le jour sans eux.

Je tiens aussi à remercier les rapporteurs et jurys de ma thèse pour leurs remarques pertinentes et l'appréciation objective de mon travail.

De manière beaucoup moins formelle, mais avec tout autant de coeur, je voudrais remercier :

– le bureau B159 : Mathias, parce que c'était un des plus vieux habitants du bureau, merci pour ta constante bonne humeur et ton soutien inconditionnel lors de nos dérives sur Koreus ou XBlast ; JP pour ton humour décalé et tes réflexions *à 2 balles*, Mathieu pour la constance et la régularité de tes interventions écologiques.
– les doctorants du pôle RIA, actuels (Julien, Benoît, Adrien, Nizar, Hung, Assia, Ali, Redouane, Naveed, Xavier, Mokhtar, Yi, Jim, Mario, Dora, Jean-Philippe, Cyril, Séverin, Akin, Arnaud, ...) et passés (Luis, Joan, Brice, ...), compagnons de travail ou de table avec qui j'ai partagé, au moins une fois, expériences et discussions.

Passons maintenant aux choses sérieuses. Un énorme merci à ma Lilie qui m'a soutenu[1] pendant toute cette aventure. Sans elle, les quelques pages qui suivent n'auraient sûrement jamais vu le jour. Elle mérite amplement une partie des honneurs, voir même un diplôme à elle tout seule.

Merci aussi à ma famille (papa, maman, mamies, Muriel et Teddy) et belle-famille (Alain, Martine, Muriel, Bernard, tata Jacqueline, ...) de m'avoir soutenu et d'avoir supporté mes changements d'humeur. Merci aussi d'avoir fait semblant de comprendre quand je vous expliquais mon travail ; maintenant, c'est fini. J'ai aussi une pensée pour les personnes qui ne sont plus là et qui, je l'espère, sont fières de moi.

Merci à Mickaël et Aurore (et Joan), Guillaume et Sandra, Matthieu, Jean-Yves et Claire, Julien, mais aussi Cédric, Pierre, Jérôme et Aurélie (et Clément) de m'avoir apporté votre amitié

[1] supporté

et de ne pas m'avoir tenu rigueur de ma non communication ces derniers mois. Les repas, matches de l'USAP, bières et autres parties de belotte et de Wii ont été autant de divertissements nécessaires durant ces trois dernières années et, je l'espère, continueront à l'être.

J'aimerai aussi remercier l'ensemble des enseignants d'IMERIR qui m'ont permis de faire ce que j'ai fait et d'être ce que je suis maintenant. Merci à Ahmed, Blaise, Eric, Gilles, Martine, Pierre et les autres, pour leur soutien et leurs conseils depuis 2003 (7 ans déjà !).

Toutes ces personnes ont, plus ou moins directement, contribué à battir et à me faire toucher du doigt mon rêve de toujours qui, je l'espère, ne va pas s'arrêter là : *"inventer des inventions"*.

Plus on avance dans la vie, plus on est obligé d'admettre
que le sel de l'existence est essentiellement dans le poivre qu'on y met.
ALPHONSE ALLAIS

Résumé

Ces travaux de thèse s'inscrivent dans le cadre du projet européen CommRob impliquant des partenaires académiques et industriels. Le but du projet est la conception d'un robot compagnon évoluant en milieu structuré, dynamique et fortement encombré par la présence d'autres agents partageant l'espace (autres robots, humains). Dans ce cadre, notre contribution porte plus spécifiquement sur la perception multimodale des usagers du robot (utilisateur et passants). La perception multimodale porte sur le développement et l'intégration de fonctions perceptuelles pour la détection, l'identification de personnes et l'analyse spatio-temporelle de leurs déplacements afin de communiquer avec le robot. La détection proximale des usagers du robot s'appuie sur une perception multimodale couplant des données hétérogènes issues de différents capteurs. Les humains détectés puis reconnus sont alors suivis dans le flot vidéo délivré par une caméra embarquée afin d'en interpréter leurs déplacements.

Une première contribution réside dans la mise en place de fonctions de détection et d'identification de personnes depuis un robot mobile.

Une deuxième contribution concerne l'analyse spatio-temporelle de ces percepts pour le suivi de l'utilisateur dans un premier temps, de l'ensemble des personnes situées aux alentours du robot dans un deuxième temps. Enfin, dans le sens des exigences de la robotique, la thèse comporte deux volets : un volet formel et algorithmique qui tire pertinence et validation d'un fort volet expérimental et intégratif. Ces développements s'appuient sur notre plateforme Rackham et celle mise en œuvre durant le projet CommRob.

Table des matières

Chapitre 1

Introduction

Ce chapitre a pour but d'introduire le contexte général de ces travaux et plus précisément ce qui touche à la perception de l'homme pendant l'interaction Homme / Robot. Nous allons présenter les enjeux de la perception de l'Homme dans le domaine de la robotique de service, ainsi que les principaux robots interactifs. Par la suite, nous allons décrire le contexte du projet CommRob dans lequel s'inscrivent ces travaux de thèse. Enfin, nous allons introduire nos travaux sur la perception embarquée de l'Homme au sein du projet CommRob.

(1) Interaction *(n.f.) :* *Relation existant entre deux éléments d'un système et qui fait que l'activité de l'un est déterminée par l'activité de l'autre.*

(2) Perception *(n.f.) :* *Action, fonction par laquelle l'esprit se représente les objets.*

1.1 Contexte et problématique

L'idée de disposer de robots en tant qu'assistants dans un lieu public ou en tant que compagnons à domicile n'est pas nouvelle ou originale. Néanmoins, les défis associés restent ouverts car les robots, une fois sortis de leurs laboratoires, doivent acquérir de nombreuses compétences sociales afin d'améliorer leur interaction avec un utilisateur novice en milieu humain encombré tel qu'une exposition, un musée ou un supermarché. Faire évoluer un robot mobile développé au sein d'un laboratoire en un robot assistant ou personnel utilisable par la majeure partie d'entre nous reste à ce jour un vrai challenge robotique. Les améliorations récentes en planification de mouvements, vision par ordinateur, traitement du langage naturel, raisonnement automatique et intelligence artificielle sont autant de jalons indispensables pour atteindre ce but.

Les récents développements en robotique et en interaction Homme / Robot ont ouvert des perspectives d'applications dans lesquelles les robots mobiles avancés apportent une assistance directe aux hommes dans leur quotidien. L'espérance de vie étant de plus en plus importante, il existe un marché potentiel réel pour de nouveaux outils robotiques. Les robots mobiles n'ont commencé que récemment à interagir avec l'homme, par exemple en s'occupant des tâches

ménagères. Leur acceptabilité par l'homme est un point essentiel et passe notamment par une réactivité du robot donc le respect de contraintes temporelles compatibles avec la tâche d'interaction. De plus, de tels robots n'ont pas vocation à remplacer l'homme au quotidien, mais à l'épauler dans ses tâches et donc de travailler en coordination avec ce dernier.

Les enjeux de la robotique interactive sont donc multiples :
- les robots assistants ou auxiliaires de services en milieu public ou professionnel requièrent des capacités définies *a priori* et liées aux services à assurer dans certaines conditions. Ils possèdent *a priori* un nombre de tâches à accomplir figées selon les services assurés.
- les robots personnels ou compagnons sont destinés à des interactions et tâches plus exclusives avec l'homme dans un lieu privatif ou à domicile. Ces derniers sont alors supposés effectuer un nombre de tâches évolutif par apprentissage en interaction avec l'homme *i.e.* accroître leurs capacités en interaction avec l'homme.

Malgré cet intérêt grandissant pour la robotique de service ainsi que les progrès conjoints de l'électronique et de l'informatique, l'introduction de la robotique d'assistance et personnelle dans la vie courante reste faible. Le robot assistant constitue encore aujourd'hui un défi de recherche très souvent ouvert sur le plan scientifique, bien que certains succès récents aient popularisé le robot mobile auprès du grand public. Nous présenterons ci-après une liste non exhaustive des réalisations en termes de projets ou plateformes significatifs développés ces dernières années.

En ce qui concerne le **robot personnel**, nombreux sont ceux qui ont tenté d'introduire des robots en tant que compagnons capables d'interagir avec une personne dans la vie de tous les jours, et ce avec plus ou moins de succès. PaPeRo (pour *Partner Personal Robot*), développé par NEC, et plus récemment Matilda, apportent une compagnie à l'homme et montre quelques aptitudes sociales rudimentaires. Ce sont des robots personnels, développés pour réaliser certaines tâches élémentaires de la maison. PaPeRo se distingue par un nombre important de capteurs. En effet, il possède différents types de capteurs lui permettant (i) de voir et de reconnaître avec ses deux caméras, (ii) d'entendre et de comprendre avec ses multiples microphones, (iii) de ressentir avec ses capteurs tactiles situés sur la tête. Entre autre, il est capable de reconnaître plusieurs personnes par l'apprentissage en ligne de différents visages. De plus, PaPeRo est capable de se déplacer dans un environnement humain grâce à des capteurs tactiles et ultrasons lui permettant de détecter les différents obstacles. Ses capacités sensori-motrices lui permettent d'accomplir des tâches telles que la surveillance d'une maison ou l'aide aux personnes âgées.

Le robot Biron [Maas et al., 2006], développé dans le cadre du projet COGNIRON, bénéficie de capacités d'interaction multimodale avancées lui permettant de détecter et de suivre les personnes dans son champ de vision mais aussi de focaliser son attention sur son utilisateur courant. Ce robot est plutôt dédié à des applications personnelles du fait qu'il n'est pas capable de différencier deux personnes. Le robot Biron intègre des capacités perceptuelles de l'environnement pour : (1) se localiser dans son environnement, (2) reconnaître la pièce dans laquelle il se trouve, (3) reconnaître les objets d'une pièce. Cependant, Biron n'a aucune capacité de navigation en présence de foules.

Récemment, Mitsubishi a développé Wakamaru [Harte and Jarvis, 2007], un robot personnel "attentionné", capable de tenir un dialogue et de reconnaître un visage humain. Ce robot a un

visage expressif et utilise des techniques de reconnaissance vocale pour interagir. A la manière de PaPeRo et Matilda, il embarque des fonctionnalités d'interaction lui permettant de détecter et de reconnaître une dizaine de personnes différentes.

Depuis 2008, le projet CompanionAble a pour but de combiner la robotique mobile avec l'intelligence ambiante afin d'aider les personnes âgées atteintes de troubles cognitifs. Pour cela, une plateforme mobile permet de monitorer les activités de son environnement afin d'en détecter d'eventuels changement inhabituels. Le projet est actuellement en cours et, pour le moment, peu d'informations sur la plateforme sont disponibles. Néanmoins, une des fonctionnalités intéressantes de cette plateforme concerne la capacité à garder l'utilisateur en contact avec son entourage plus ou moins direct (amis, relations, medecins, etc) grâce à l'utilisation de la visiophonie. Pour cela, le robot doit observer la personne au moyen de caméras afin (i) d'en extraire l'information de l'image d'entrée, (ii) de relayer ces données vers la personne concernée [Schroeter et al., 2009].

Au delà des projets et systèmes présentés, la plupart des robots personnels se basent sur une ***interaction active*** pour échanger avec l'utilisateur, au détriment parfois d'une compréhension globale de l'environnement. En effet, de tels robots doivent agir de manière réactive aussi bien avec l'homme qu'avec l'environnement. Or, l'intégration de ces capacités sur un robot personnel reste un problème ouvert en robotique mobile, et, à notre connaissance, très peu de robots personnels sont capables de réaliser un nombre conséquent de tâches.

Concernant les **robots assistants**, de nombreux projets ont permis d'accomplir des avancées significatives dans le domaine de l'interaction Homme / Robot en environnement public. A notre connaissance, Rhino [Burgard et al., 1998] fut le premier robot à être déployé dans un musée. La seconde génération de robots, comme Minerva, a ensuite suivi cet exemple [Thrun et al., 2000]. Bien évidemment, Minerva surpasse Rhino dans bien des domaines. En effet, il utilise un algorithme probabiliste d'apprentissage de cartes ainsi qu'un ensemble accru de capacités interactives. La prise en compte de la position de l'utilisateur lors de l'interaction se fait au travers de données laser. Tout comme Minerva, Tourbot [Trahanias et al., 2000], Mobot [Nourbakhsh et al., 2003] et Cicerobot [Chella et al., 2007] sont capables de fournir une analyse précise de leur environnement, mais n'intègrent que des capacités perceptuelles limitées sur l'homme ce qui limite leur interaction.

Un autre robot assistant a été développé au sein du projet MORPHA [Bischoff et al., 2002]. Ce projet se focalise sur différents aspects de l'interaction Homme / Robot, comme l'interaction multimodale, la compréhension de situations Homme / Robot, l'identification de l'utilisateur et l'interprétation de ses intentions, ainsi que la coordination des mouvements et des actions entre le robot et son utilisateur. Même si MORPHA a amené de remarquables résultats, principalement dans l'execution de tâches de manipulation, le robot prototypé opère et interagit uniquement dans un environnement simple tel une maison pour les tâches quotidiennes ou dans un atelier pour des applications industrielles. Dans ces deux cas, seuls des mouvements locaux au sein d'un environnement statique et connu sont possibles, avec pour cela, une communication exclusive entre le robot et son utilisateur.

Le démonstrateur mobile Robox [Siegwart et al., 2003] tient une place importante dans le do-

maine du robot d'assistance du fait de son déploiement à large échelle pour éprouver ses capacités de navigation et d'interaction. Robox utilise différents capteurs (*i.e.* laser, caméra, microphone) et des algorithmes dédiés afin de détecter les visiteurs de l'Expo.02 en Suisse. Onze Robox ont été déployés pendant les 159 jours de l'exposition. Ces expérimentations ont montré que 70% des personnes ayant utilisées Robox seraient favorables à une utilisation plus régulière dans un supermarché ou une gare, par exemple.

Alpha [Bennewitz et al., 2005] est capable d'interagir avec plusieurs personnes par l'analyse du flux audio-visuel lui permettant de basculer d'un interlocuteur à l'autre. Cette approche fonctionne correctement dans un environnement faiblement perturbé, mais peut difficilement gérer les situations fortement encombrées. De plus, bien que conçu comme un humanoïde, ses capacités de déplacement dans un environnement humain restent limitées.

La plupart des systèmes pré-cités [Thrun et al., 2000; Siegwart et al., 2003; Maas et al., 2006] concentrent aussi leur approche sur des détections laser afin de détecter et suivre des personnes aux alentours du robot, tout comme Hermes [Bischoff and Graefe, 2004], Maggie [Gorostiza et al., 2006] et Pearl [Pineau et al., 2003]. Les données laser sont aussi largement utilisées lors de la navigation d'un robot mobile afin d'éviter les obstacles de manière dynamique. En plus des détections laser, Robovie [Kanda et al., 2004] utilise de nombreux capteurs embarqués (caméra, microphone), mais aussi déportés dans l'environnement (lecteur de badges RFID, capteurs de pression) afin de localiser et d'identifier les personnes présentes aux alentours. Ces systèmes dédiés à la robotique d'assistance sont alors capables de combiner une ***interaction active*** avec un utilisateur à l'instar d'un robot personnel ainsi qu'une ***interaction passive*** leur permettant de partager leur environnement avec les passants afin de naviguer facilement dans ce milieu humain donc *a priori* très dynamique.

Les travaux présentés ci-après se focalisent principalement sur la perception embarquée de l'homme pour des robots assistants. En effet, les capacités perceptuelles d'un robot personnel sont assez classiques car le robot n'interagit qu'avec une seule personne, alors qu'un robot assistant doit développer une perception plus complexe des humains du fait qu'il évolue dans un milieu encombré.

Chacun des systèmes présentés met en avant des fonctionnalités perceptuelles spécifiques, très adaptées au contexte d'application restreint. Bien que la plupart aient donnés des résultats probants, leurs capacités de perception visuelle de l'homme restent limitées et, malgrè leur intérêt, les plateformes fusionnant des données issues de différents capteurs ne sont que très faiblement représentées. De plus, l'interaction Homme / Robot est souvent découplée de la partie navigation et très peu de systèmes gèrent les mouvements des humains à la fois pour suivre une personne cible et éviter les personnes obstacles de façon intelligente.

1.2 Description du projet CommRob

Cette thèse s'inscrit, notamment par son financement associé, dans le projet européen STREP CommRob[1], un projet financé par la division FP6 de la Commission Européenne. Le projet CommRob (pour *Communication with and among robots*) comporte cinq partenaires académiques ou industriels *i.e.* FZI (Karlsruhe, Allemagne), KTH (Stockholm, Suède), LAAS (Toulouse, France), TUW (Vienne, Autriche), Zenon (Athènes, Grèce).

1.2.1 Contexte du projet

Ce projet vise à développer un robot assistant capable d'évoluer en environnement humain très encombré. Il a pour but de mettre en avant les avancées dans le domaine de la communication de haut niveau avec / entre robots. Au delà de la communication par la parole, la communication multimodale considère différents percepts (déplacements, gestes, etc) afin de définir une interaction Homme / Robot plus riche. La mise en place d'une approche unifiée de la communication est un des objectifs majeur du projet. Cette approche avancée permet de faciliter indépendamment la proche coopération entre l'homme et le robot ou entre robots. L'échange d'information entre robots concernant l'état de l'environnement mutuellement partagé permet de rendre la navigation en zone encombrée plus efficace et sûre.

De nombreux problèmes liés aux comportements avancés des robots en environnements dynamiques sont abordés *i.e.* la localisation basée sur des amers, l'apprentissage de carte topologiques indexées par ces amers, la navigation autonome ainsi que la perception de l'utilisateur et la détection et l'évitement d'obstacles dynamiques. Dans ce but, un autre objectif majeur du projet est alors d'aborder des environnements plus complexes, car densemment peuplés, pour les robots par rapport aux travaux cités précédemment. Un autre challenge est de faire cohabiter deux tâches de navigation réactives, *i.e.* le suivi d'une personne cible et l'évitement des passants.

Ce projet vise à la fois le prototypage d'un tel robot et des évaluations poussées dans le contexte applicatif puisque ce robot est destiné à servir de chariot dans un supermarché, un aéroport, etc. En plus de pouvoir transporter des produits, ce robot doit être capable de guider un utilisateur à travers un environnement structuré, dynamique et complexe (*i.e.* interaction active) tout en évitant les personnes partageant l'espace (*i.e.* interaction passive).

1.2.2 Scénarii visés au sein du projet

Différents scénarii et modes d'utilisations sont définis afin de mettre en avant les différentes fonctionnalités développées au sein du projet ainsi que leur intégration. Le **scénario S1** et le **scénario S2** n'impliquent pas notre problèmatique de perception de l'homme. Ils correspondent aux scénarii décrits ci-après mais *via* un contact physique entre le robot et son utilisateur par une poignée haptique. Nous listons ici les scénarii et modes relatifs à notre travail, bien que d'autres scénarii complémentaires aient été définis.

[1] voir le lien URL `www.commrob.eu`.

– **Scénario S3**
  Ce scénario est défini pour une personne habituée au supermarché. Le robot n'a donc pas l'initiative des mouvements. En effet, l'utilisateur agit comme agent maître. Les déplacements du robot se font sans contact, grâce à la perception multimodale de l'utilisateur. Le robot accompagne alors l'utilisateur en respectant un certaine distance de sécurité, tout en détectant et évitant les obstacles. Le robot avertit l'utilisateur du moindre problème, *e.g.* l'utilisateur avance trop vite ou un obstacle est trop encombrant, par une interface vocale. De même, le robot s'arrête lorsque l'utilisateur s'arrête. Les fonctionnalités mises en œuvre dans ce scénario sont les suivantes :
  – identification de l'utilisateur par radio fréquence et vision,
  – suivi multimodal de l'utilisateur afin de s'assurer de sa présence durant les tâches de mouvements coordonnés.

  Ce scénario permet de définir un premier *mode d'interaction* entre le robot et l'homme : ***le suivi*** de l'homme par le robot (ou *following mode*).
– **Scénario S4**
  Ce scénario est défini pour une personne ne connaissant pas le supermarché ou nécessitant une assistance. Le robot a donc l'initiative des mouvements. Comme précédemment, l'interaction entre l'homme et le robot se fait sans aucun contact physique. Le robot doit alors s'assurer de la présence de l'utilisateur. En effet, une certaine distance doit être conservée entre le robot et l'homme. Les fonctionnalités de perception de l'utilisateur nécessaires à l'exécution de ce scénario sont identiques au scénario précédent. Ce scénario permet de définir un autre *mode d'interaction* entre le robot et l'homme, *a priori* plus facile au niveau perceptuel car l'homme fait face au robot : ***le guidage*** de l'homme par le robot (ou *guiding mode*).

Dans cette perspective, deux modes d'interaction sont envisagés, une ***interaction active*** et une ***interaction passive***, afin de guider / suivre une personne cible tout en évitant les passants de manière naturelle est sociable. Le choix du *mode d'interaction* (suivi ou guidage) est défini par l'utilisateur lors de l'initialisation de l'interaction avec le robot alors que les deux *types d'interaction* (active ou passive) cohabitent en permanence et sont définis en fonction des personnes mises en jeu : l'utilisateur pour la perception active, les passants pour la perception passive.

1.2.3 Enjeux et problématique de la perception embarquée de l'Homme au sein du projet

Nos travaux visent à caractériser les relations dynamiques Homme / Robot impliquées dans les différents scénarii et modes pré-cités. Chaque interaction commence lorsque l'utilisateur focalise son attention sur le robot. Pendant l'execution d'un scénario (ou mission), le robot doit s'assurer de la présence et de l'identité de l'utilisateur. Concernant les missions de guidage ou de suivi de l'homme par le robot, ce dernier doit non seulement être capable de suivre de manière robuste l'utilisateur, mais aussi de détecter les possibles erreurs dues à la perte de la cible ou à son occultation. La conception d'algorithmes basés sur la vision doit être assez efficace et robuste aux occultations temporaires de la cible, aux situations encombrées et aux changements

des conditions d'illumination.

Le suivi, ou analyse spatio-temporelle, constitue une fonctionnalité perceptuelle clé lorsqu'il s'agit de caractériser, à partir de données capteurs, la relation d'un robot mobile à des personnes *a priori* mobiles. En milieu humain encombré, un traqueur doit s'appuyer sur des détecteurs robustes et identifier les personnes afin de s'affranchir du problème d'association de données lors du suivi.

Ces considérations entraînent le développement de fonctions qui peuvent être catégorisées comme suit :

- **Les fonctions de détection/reconnaissance de personnes** : différentes méthodes de détection de personnes issues de différents capteurs (caméras, laser, etc) sont étudiées. Des méthodes de (ré-)identification sont proposées ainsi que leur paramétrage afin de quantifier leur apport dans le suivi et la récupération de cibles.
- **Le suivi multi-sensoriel de personnes** : les méthodes de suivi sont implémentées afin d'évaluer la position de la ou des personnes dans le repère image ou robot. Un enjeu est d'obtenir un suivi de personne(s) robuste (aux encombrements et occultations), quitte à être moins précis, utilisant principalement la vision lorsque les personnes sont à quelques mètres ou plus précisément à distance sociale du robot.

Ces fonctionnalités, sont rattachées au *Work Package 4* (WP4) du projet : *Human Motion Interpretation*. Le but du *WP4* consiste à implémenter des fonctions pour détecter, reconnaître, suivre des personnes et à définir les modalités basiques de l'interaction multimodale Homme / Robot basées sur l'observation et l'analyse de l'homme. Le robot est sensé évoluer dans un environnement hautement dynamique et encombré alors que la puissance de calcul embarqué est souvent limitée. L'accent est mis sur le développement d'algorithmes efficaces et robustes aux artefacts de l'environnement *e.g.* les occultations. Plusieurs flots de données sensorielles doivent être gérées simultanément dans les algorithmes développés et une intégration robuste et probabiliste de ces différents percepts est à envisager.

1.2.4 Fonctionnalités perceptuelles et projet CommRob

Le WP4 est découpé en différentes tâches. Les tâches, **T4.1** relative aux spécifications fonctionnelles et **T4.2** relative au développement d'une interface haptique pour la commande du robot, ne sont pas détaillées car elles sortent du cadre de nos travaux. Nous nous focalisons ci-après sur les fonctionnalités développées durant cette thèse *i.e.* :

T4.3 : Fonctionnalités pour la détection et l'identification de personnes

Un ensemble de fonctions bas niveau est défini pour la détection et l'identification de personnes. Des primitives visuelles sont disponibles pour détecter les différentes personnes présentes dans le champ de vue du robot. Ces fonctionnalités et leurs attributs associés peuvent aussi être utilisés pour le suivi visuel, autant dans la procédure d'initialisation que comme source d'information dans la boucle principale de suivi. En plus de la détection, l'identité de l'utilisateur doit aussi être vérifiée tout au long de l'interaction. Ceci implique la définition de fonctionnalités

reposant sur (1) des attributs visuels basés sur des informations générales comme la couleur, les contours, les mouvements, (2) un badge RFID porté par l'utilisateur pour son identification, (3) des mesures de distances issues d'un laser.

T4.4 : Suivi de personnes basé sur la vision pour les mouvements coordonnés Homme / Robot

Le but est de suivre l'utilisateur à différentes distances relatives *i.e.* proximale (*e.g.* lorsque l'utilisateur interagit avec l'interface du robot), sociale (*e.g.* pendant une mission de guidage), distante (*e.g.* lorsque la personne est reconnue comme étant l'utilisateur, afin de s'approcher de l'utilisateur). Le traqueur développé se base alors sur les attributs bas niveau définis dans **T4.3**. L'estimation de la position de l'utilisateur servira à commander un déplacement du robot par rapport à l'utilisateur de manière sociale. Il faut donc savoir faire face à des situations exceptionnelles, comme un échec du suivi. Une vision active est alors requise et sera donc considérée grâce à l'utilisation d'une platine orientable. Dans la même veine, une extension à un traqueur multi-personnes pour l'interaction passive est ensuite proposée.

A l'instar des tâches T4.1 et T4.2, la tâche **T4.5** ne concerne pas directement nos travaux, bien qu'elle soit relative à l'interaction Homme / Robot. En effet, la tâche T4.5 vise à interpréter des commandes envoyées par l'homme au robot par l'intermédiaire de commandes multimodales fusionnant gestes et parole. Le suivi et l'interprétation des gestes, la reconnaissance d'ordres vocaux ainsi que la fusion de ces deux canaux de communication ont été abordés par B. Burger dans sa thèse [Burger, 2010].

1.3 Objectifs de la thèse

De par les spécifications générales et propres au projet CommRob, l'objectif de nos travaux est de proposer des stratégies de fusion de données pour la perception multimodale de l'homme en environnement encombré. Nos travaux proposent une interface perceptuelle ayant pour but (1) d'établir de manière robuste et continuelle un contact visuel entre le robot et son utilisateur, (2) de percevoir les personnes environnantes. L'ensemble des fonctionnalités nécessaires à ce robot assistant est schématisé par le synoptique de la figure 1.1.

Listons ci-après les différentes fonctionnalités mises en jeu et détaillées dans les prochains chapitres, ainsi que leurs spécificités :

1. **Détection, localisation et identification** : Le but est de fournir un ensemble de primitives issues des différents capteurs visuels, RF ou encore laser. Chaque mesure donne une information plus ou moins précise sur l'identité ou la position des personnes présentes autour du robot. L'aspect plus ou moins intermittent de chaque mesure donne une information fiable qui doit s'inscrire dans une analyse spatio-temporelle. Les travaux sur la détection, l'identification et la localisation de personnes à partir de microphones qui ne seront pas traîtés ici font actuellement l'objet d'une thèse associée [Bonnal et al., 2009].
2. **Fusion multimodale de données hétérogènes** et **suivi mono-cible** : Le but est de mixer les différents attributs extraits au niveau sensoriel de l'architecture. Le suivi utilise un modèle

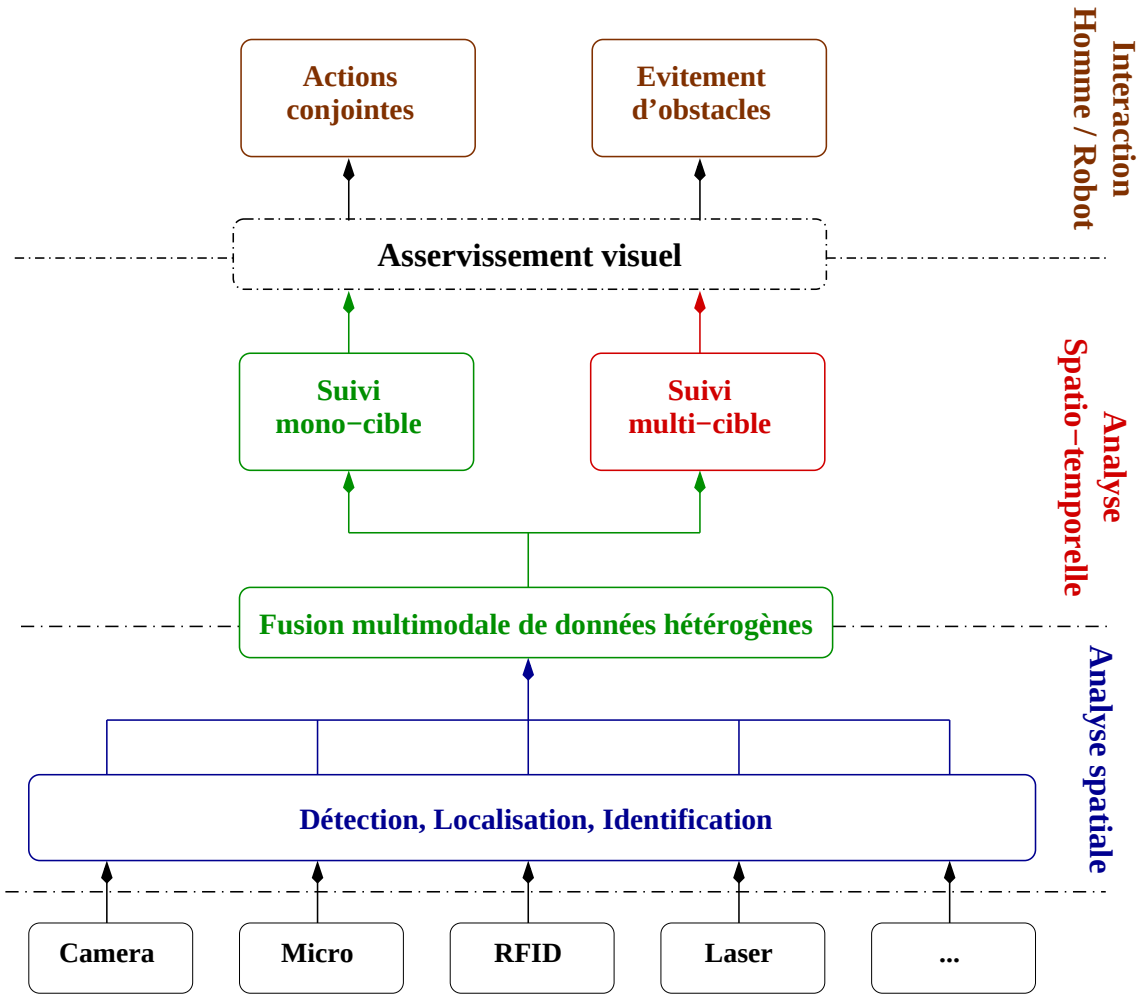


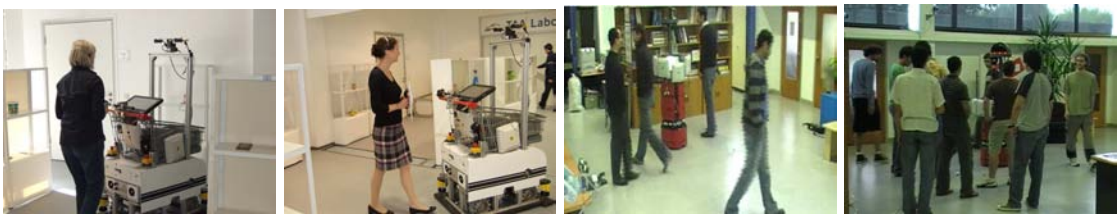
FIG. 1.1 – Synoptique de notre système de perception multimodale de l'homme.

précis de l'identité de l'utilisateur pour estimer sa présence ou non dans le champ de vue du robot. Il s'agit ici d'estimer la position de l'utilisateur par des méthodes stochastiques permettant de palier à l'aspect sporadique des détecteurs et de gérer avec robustesse les artefacts de l'environnement *i.e.* les occultations, la variabilité de l'apparence de la cible, l'évolutivité de l'environnement.

3. **Suivi multi-cibles** : il permet d'estimer les déplacements des passants au voisinage immédiat du robot. Les données issues de différents capteurs sont utilisées afin de suivre chaque personne de manière fiable.
4. **Actions conjointes** : l'asservissement des déplacements du robot est réalisé grâce aux résultats du suivi mono-cible. Au travers d'un asservissement visuel, le robot effectue des déplacements en relations avec ceux de l'utilisateur afin de réaliser une tâche de guidage

ou de suivi en restant à une distance sociale de ce dernier. L'**évitement d'obstacles** utilise, lui, les résultats du suivi multi-cibles pour évaluer la trajectoire idéale permettant de réaliser une tâche robotique conjointe. Le robot évolue donc en fonction des personnes environnantes afin de ne pas perturber leurs déplacements. La figure 1.2 illustre quelques situations types Homme / Robot à gérer.

L'articulation de nos travaux réside donc dans un couplage fort entre la perception de l'homme et la commande des actionneurs du robot (moteurs, platine) dans le but (i) de s'asservir sur les déplacements de l'utilisateur (en interaction active) et des passants (en interaction passive), (ii) d'orienter la caméra pour diriger les ressources perceptuelles vers les zones pertinents de l'espace environnant. Les travaux relatifs à la commande du robot sont traîtés en forte collaboration avec d'autres membres du groupe et ont donné lieu à deux thèses associées [Ouadah et al., 2009; Durand-Petiteville et al., 2010].



(a) (b) (c) (d)

FIG. 1.2 – Situations Homme / Robot impliquant un déplacement conjoint entre le robot et son utilisateur ((a) et (b)), ainsi que l'évitement des personnes de la scène ((c) et (d)). Les situations (a) et (b) impliquent la plateforme de CommRob, Inbot, alors que les situations (c) et (d) impliquent notre plateforme Rackham.

1.4 Plan du document

Ce chapitre a présenté un grand nombre des plateformes robotiques interactives existantes afin de positionner notre système global par rapport à l'état de l'art. L'organisation de ce manuscrit suit l'organisation de nos travaux au sein du projet CommRob en décrivant successivement les fonctionnalités listées ci-dessus. Chaque chapitre traite d'un volet spécifique de nos travaux et contient l'étude bibliographique et les évaluations qui lui sont associées.

La détection et l'identification de personne est décrite dans le chapitre 2. Ce chapitre présentera deux contributions significatives *i.e.* l'identification visuelle de visages et l'identification RFID par l'adaptation d'un système du commerce, mais aussi des détecteurs de personnes relatifs à la littérature, mais utiles à la construction de notre approche.

La fusion de données hétérogènes ainsi que le suivi mono-cible sera présenté dans le chapitre 3 afin de mettre en avant une **approche percetuelle active** et exclusive de l'utilisateur. Ce chapitre se base sur les fonctionnalités présentées au chapitre précédent. Des évaluations qualitatives et quantitatives sont détaillées afin de valider l'approche proposée.

Nos travaux préliminaire sur le suivi multi-cibles seront détaillés dans le chapitre 4 et permettront de définir les bases d'une **interaction passive** entre le robot et les personnes présentes dans son environnement direct. Ici aussi, les primitives définies au chapitre 2 seront utilisées dans le cadre d'une fusion de données multi-capteurs. Des évaluations préliminaires seront présentées en fin de chapitre.

Le chapitre 5 porte lui sur l'intégration de nos fonctionnalités sur différentes plateformes robotiques. La réalisation des scénarii S3 et S4 présentés en section 1.2 et l'évaluation des fonctionnalités relatives aux déplacements conjoints Homme / Robot et à l'évitement de personnes seront détaillés lors d'expérimentations robotiques.

Enfin, le chapitre 6 résume l'ensemble de nos contributions et présentes les perspectives associées à ces travaux.

Chapitre 2

Détection et identification de personnes

Dans un processus d'interaction Homme / Robot, il est indispensable de pouvoir localiser et identifier l'interlocuteur privilégié du robot. En effet, avant d'être autorisée à interagir avec le robot, chaque personne doit être identifiée afin de s'assurer de la cohérence des échanges. Ce chapitre traite de la détection et de l'identification de personnes au travers de données issues de différents capteurs. Détection et identification sont indispensables lors de la réalisation d'une tâche robotique, et ceci à deux niveaux.

Tout d'abord, au niveau fonctionnel, une personne interagissant avec un robot doit être identifier de manière sûre et fiable par ce même robot. Ceci implique de pouvoir obtenir une signature précise de l'utilisateur. De plus, dans notre contexte applicatif, il est nécessaire qu'une personne soit identifiée dès les premiers instants de l'interaction, *e.g.* un apprentissage en ligne ne nécessitant aucun post-traitement de la part d'un opérateur externe. Pour ce faire, il existe de nombreuses approches de détection et d'identification de personnes qui diffèrent notamment par le type de capteurs employés.

Au niveau perceptuel, la détection et l'identification des différentes cibles susceptibles d'intéragir avec le robot sont vitales pour s'affranchir des problèmes d'association de données. En effet, certains capteurs comme le laser ou les infrarouges sont très fiables lorsqu'il s'agit de détecter une personne, mais pêchent à différencier deux cibles du fait que l'information délivrée est pauvre. Nous nous proposons d'évaluer différentes méthodes permettant d'identifier avec certitude une cible *via* différents percepts sensoriels. En effet, il est important de jouer sur la complémentarité des capteurs mis en jeu.

Ce chapitre est structuré comme suit.

La section 2.1 présente notre processus d'identification visuelle de visages. Après une présentation générale de la problématique et un état de l'art des techniques d'identification visuelle, nous présentons notre propre système. Les performances relatives aux différents classifieurs issus de la littérature sont étudiées et analysées afin de définir le meilleur jeu de paramètres associés pour chaque classifieur. Ceci constitue ici la spécificité de nos travaux. Par la suite, les paramètres libres nécessaires à l'exécution du classifieur proposé sont optimisés afin d'obtenir de meilleures performances.

La section 2.2 présente l'adaptation d'un capteur RFID du marché permettant la détection et

l'identification d'une personne au voisinage du robot sur $360°$. De même que pour l'identification visuelle, nous présentons tout d'abord quelques généralités sur la problématique d'identification Radio Fréquence (RF) ainsi qu'un état de l'art des techniques utilisées dans la littérature. Ensuite, notre capteur RFID est décrit puis évalué dans notre contexte applicatif. Les investigations en cours visant à améliorer la compacité du système RFID embarqué sont présentées.

Des besoins récents et en marge de la plateforme de CommRob, Inbot, ont nécessité la mise en œuvre de détecteurs laser et visuel mais non restreints au visage. Ces détecteurs sont présentés dans la section 2.3 *i.e.* les détections basées sur le laser et la détection visuelle de personnes.

La section 2.4 conclue alors sur les contributions de ce chapitre ainsi que les perspectives relatives à la détection et l'identification de personnes.

2.1 Identification visuelle de visages

2.1.1 Considérations générales

La reconnaissance visuelle de personnes depuis une plateforme mobile opérant dans un environnement humain est un défi qui impose différentes contraintes. Tout d'abord, la puissance de calcul embarquée sur le robot doit permettre l'exécution concurrente et complémentaire d'autres fonctionnalités non visuelles ainsi que de routines décisionnelles au sein de l'architecture logicielle du robot. De plus, une attention particulière doit être apportée à la conception d'algorithmes robustes aux conditions environnementales variables. Au contraire des systèmes biométriques conventionnels, le capteur de vision embarqué évolue dans un environnement humain non nécessairement coopératif où les personnes se tiennent à quelques mètres – approximativement à distance sociale ($[1.2, 3.5]$m) ou personnelle ($[0.5, 1.2]$m) [Hall, 1966] – au moment de l'interaction avec le robot.

Dans ce contexte, notre système de reconnaissance de visage doit être capable de gérer (i) une image de faible résolution permettant des fréquences d'acquisition plus grandes, (ii) de forts changements des conditions d'illuminations, (iii) de larges variations dans la pose du visage (figure 2.1) tant en 2D (plan image : roulis) qu'en 3D (tangage et lacet), (iv) des occultations partielles et l'encombrement de la scène inhérents au contexte applicatif.

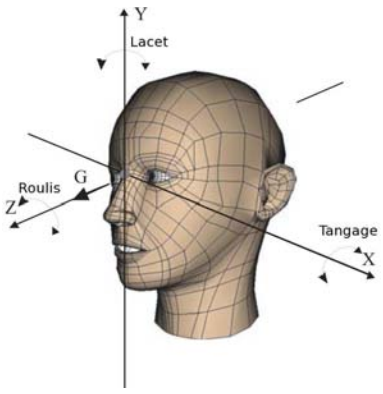


FIG. 2.1 – Schéma des notions de tangage, lacet et roulis.

Au delà des techniques basées sur des images fixes, les approches exploitant l'information spatio-temporelle ont récemment été utilisées pour le contrôle d'accès ou la vidéo-surveillance. Ces approches peuvent, elles aussi, être catégorisées comme suit : (1) les systèmes de reconnaissance de visages vidéo depuis des images fixes (ou *still-to-video*) [Choudhury et al., 1999], basées sur une galerie d'images fixes, et (2) les systèmes de reconnaissance de visages vidéo depuis une vidéo (ou *video-to-video*) [Aggarwal et al., 2004].

Ces dernières ne sont pas forcément adaptées à notre contexte robotique qui comprend de nombreux mouvements au sein même de l'image en terme de distance, de rotations 2D et 3D. L'analyse spatio-temporelle, c'est à dire le suivi, est, dans de nombreux cas, basée sur des méthodes de simulation de Monte Carlo, aussi connues sous le nom de filtres à particules [Doucet et al., 2001]. Notre étude visant à produire un détecteur fiable mais, par nature, intermittant, nous allons nous concentrer sur les approches d'identification de visages à partir d'images fixes que nous intégrerons alors dans le processus de suivi (c.f. chapitre 3), notamment pour permettre sa réinitialisation automatique.

Historiquement, l'identification vidéo de visages est issue des techniques de reconnaissance depuis des images fixes. En d'autre termes, ces systèmes détectent et segmentent automatiquement un visage depuis une vidéo et utilisent ensuite des techniques appliquées aux images fixes. Généralement, ces méthodes peuvent être classées dans deux catégories : (1) les méthodes dites holistiques ou globales, et (2) celles dites locales *i.e.* basées sur des caractéristiques locales du visage, même si quelquefois, les deux approches sont combinées au sein d'un même système [Lam and Yan, 98]. Bien évidemment, dans notre contexte robotique, une méthode locale (ou *feature-based*) [Quintiliano et al., 2001] n'est pas vraiment adaptée. En effet, de petites images de visages (dues à la distance Homme / Robot) ainsi qu'une faible résolution de l'image acquise complique l'extraction de caractéristiques faciales sur un visage. De plus, les approches holistiques (ou basées apparence) [Belhumeur et al., 1996; Shan et al., 2003; Turk and Pentland, 1991] considèrent le visage comme une entité et traitent directement l'intensité de chaque pixel représentant un visage en évitant d'extraire des caractéristiques locales. Le lecteur peut aussi se référer aux études [Abata et al., 2007; Zhao et al., 2000] pour une analyse plus détaillée de l'état de l'art relatif aux techniques de reconnaissance de visage basées sur des images fixes.

Dans ce chapitre, nous mettrons donc l'accent sur le développement d'un système de reconnaissance de visages basé sur des images fixes avec contraintes temporelles.

2.1.2 Etat de l'art

Depuis les années 1990, les méthodes basées sur l'apparence ont été privilegiées dans les systèmes de reconnaissance de visage. Classiquement, elles reposent sur trois étapes séquentielles : (1) un pré-traitement des images brutes, (2) la projection d'images dans un sous-espace afin de construire une représentation de l'image dans un espace de dimension inférieure pour réduire le temps de calcul, (3) une règle de décision finale pour la classification. Adini *et al.* dans [Adini et al., 1997] mettent en exergue le rôle du prétraitement.

Une étude préliminaire sur les pré-traitements a été réalisée dans [Germa et al., 2007a]. Il s'avère, comme évoqué dans [Heseltine et al., 2002; Jonsson et al., 2000], que l'égalisation d'histogramme offre le meilleur compromis entre la rapidité de traitement et la performance quant à la mise en valeur des informations contenues dans une image.

En ce qui concerne les techniques de réduction de dimensionnalité, Pang *et al.* dans [Pang et al., 2006] utilisent des techniques de projection non-linéaires basées sur des méthodes LLE (pour *Locally linear embedding*). Celles-ci, bien que performantes, apparaissent lourdes et com-

plexes au regard de notre contexte. L'analyse en composante principale (ACP), l'analyse factorielle discriminante (AFD) ainsi que l'analyse en composante indépendante (ACI) sont des approches linéaires répandues pour la projection d'images permettant de réduire la dimension de l'espace manipulé. L'ACP utilise la projection d'images sur des *images propres* afin de déterminer les vecteurs de base qui composent l'image de variance maximale par classe [Shan et al., 2003] ou pour l'ensemble des classes [Turk and Pentland, 1991]. L'AFD détermine l'ensemble optimal de vecteurs de base assez discriminants pour que le rapport entre la dispersion inter-classe et intra-classe soit maximal. L'AFD trouve donc la meilleure base de projection dans laquelle les échantillons d'apprentissage des différentes classes sont au mieux séparés. L'AFD est indifféremment utilisée sur les images brutes afin d'en extraire les visages de Fisher [Belhumeur et al., 1996; Jonsson et al., 2000] ou combinée à l'ACP pour obtenir des attributs propres discriminants [Zhao and Chellapa, 1998]. L'ACI propose un ensemble de vecteurs de base possédant une indépendance statistique maximale [Bartlett et al., 2002]. Nos évaluations portent sur des bases de visages représentées classiquement par des sous-espaces ACP ou AFD dont le paramètre libre principal est l'énergie cumulée (notée η) des vecteurs propres.

Par la suite, il faut définir une régle de décision permettant de compléter notre système de reconnaissance. La norme Euclidienne [Jonsson et al., 2000], la distance de Hausdorff [Lin et al., 2003], la distance à l'espace de visage (ou DFFS pour *Distance From Face Space*) [Turk and Pentland, 1991] donnent de bons résultats. Lors d'études préliminaires [Germa et al., 2007a], nous avons proposé une norme d'erreur qui a donné de meilleurs résultats que la DFFS. Ces règles de décision requièrent un seuil de décision annoté τ par la suite. A l'instar de [Jonsson et al., 2000], les évaluations sont étendues aux Machines à Vecteurs de Support (ou SVM pour *Support Vector Machine*) en complément d'une ACP permettant de réduire la dimension de l'espace de représentation. Les SVM projettent une observation depuis un sous-espace d'entrée dans un espace de dimension supérieure utilisant une transformation non nécessairement linéaire, puis utilise un hyperplan de cet espace qui maximise la marge de séparation afin de minimiser le risque de mauvaise classification entre les visages.

L'optimisation automatique des différent paramètres libres de tels systèmes est souvent faite soit de manière empirique, soit au travers de courbes ROC (pour *Reciever Operator Characteristics*) [Gavrila and Munder, 2007; Provost and Fawcett, 2001], soit au moyen de méthodes numériques utilisant des fonctions non linéaires de minimisation d'objectifs. Dans cette optique, les méthodes locales de descente de gradient [Chapelle et al., 2002] ou d'optimisation globale [Boardman and Trappenberg, 2006; Yang et al., 2006] sont proposées pour améliorer les performances. Les algorithmes génétiques (ou GA pour *Genetic algorithm*) sont des techniques très connues pour les problèmes d'optimisation, et se sont avérés être vraiment efficaces pour déterminer les paramètres des SVM [Seo, 2007; Xu and Li, 2006]. L'utilisation de ces divers outils permet alors de définir les valeurs des paramètres libres nécessaires à chaque modèle de classifieur afin d'en exhiber les performances optimales de chacun. Bien qu'amenant un intérêt non négligeable pour l'optimisation des paramètres, à notre connaissance, ce point semble peu discuté dans la littérature.

2.1.3 Description de notre classifieur

Forts de ces considérations générales, nous avons réalisé des expérimentations sur la reconnaissance depuis un prétraitement basé sur l'égalisation d'histogramme, deux représentations différentes (ACP et AFD), et trois règles de décisions (norme d'erreur, distance de Mahalanobis et SVM) car ces outils sont les plus exploités dans la littérature. La figure 2.2 présente le synoptique du système. Rappelons que le but final est de classifier des visages $\mathcal{F} = (\mathcal{F}_i)_{i=1}^{nm}$ (avec $n \times m$ la résolution de l'image), extraits d'une image d'entrée, dans une des classes C_t de l'ensemble $\{C_l\}_{l=1}^{M}$ de M classes de visages à l'aide d'algorithmes d'apprentissage. Pour détecter des visages, nous utilisons le detecteur de Viola *et al.* [Viola and Jones, 2001], étendu dans [Viola and Jones, 2003; Wang et al., 2006] pour couvrir des orientations de visages de $\pm 45°$. Chaque visage extrait par le détecteur de Viola est alors utilisé en entrée du système de reconnaissance de visages.

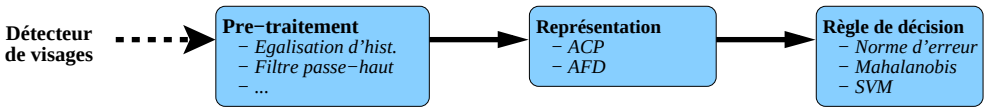


FIG. 2.2 – Processus d'identification de visages.

Sous-espace de représentation

La base de ***visages propres*** (ou *Eigenface*) W_{acp}, basée sur l'analyse en composante principale (ou ACP), est déduite en résolvant :

$$S_T.W_{acp} - W_{acp}.\Lambda = 0, \tag{2.1}$$

avec S_T la matrice de dispersion des individus $\{\mathcal{F}\}$ au sein de la classe à modéliser et Λ le vecteur des valeurs propres ordonnées. Nous ne conservons alors comme base de projection que les N_v premiers vecteurs propres tels que :

$$\frac{\sum_{i=0}^{N_v} \Lambda_i}{\sum \Lambda_i} \leq \eta, \tag{2.2}$$

avec η un ratio de la covariance totale prédéfini appelé énergie.

Une autre approche consiste à utiliser les ***visages de Fisher*** (ou *Fisherspace*) basés sur l'analyse fonctionnelle discriminante (ou AFD). La base de visages de Fisher W_{afd} est déduite de :

$$S_B.W_{afd} - S_W.W_{afd}.\Lambda = 0, \tag{2.3}$$

ou S_B et S_W sont les matrices de dispersion inter-classe et intra-classe. La sélection du nombre de vecteurs propres à conserver suit le principe défini dans l'équation 2.2.

Règle de décision

Les classifieurs de visages considèrent différentes règles de décision. Parmi elles, nous nous proposons d'en évaluer trois, à savoir (i) la norme d'erreur, (ii) la distance de Mahalanobis et (iii) les SVM.

La règle de décision basée sur ***la norme d'erreur*** introduite dans [Germa et al., 2007a] se formalise comme suit.

Soit $\mathcal{F} = (\mathcal{F}_i)_{i=1}^{nm}$ un visage à évaluer et $\mathcal{F}_{r,t}$ la reconstruction de ce visage dans la base de visage W_t relative à la classe C_t, de prototype $\mathcal{M}_t$, telle que $\mathcal{F}_{r,t} = W_t.(\mathcal{F} - \mathcal{M}_t).W_t^{'}$. La norme d'erreur est donnée par :

$$\mathscr{D}_{C_t}(\mathcal{F}) = \sum_{i=1}^{nm}(\mathcal{F}(i) - \mathcal{F}_{r,t}(i) - \mu_{\mathcal{F}})^2, \tag{2.4}$$

où $\mathcal{F} - \mathcal{F}_{r,t}$ est l'image-différence de moyenne $\mu_{\mathcal{F}}$. Cette norme d'erreur décrite dans [Germa et al., 2007a] a montré de meilleurs résultats que la norme Euclidienne et la DFFS. La figure 2.3 détaille les étapes du calcul sur trois classes différentes.

(a) $\mathcal{F}$ (b) $\mathcal{M}_t$ (c) $\mathcal{F} - \mathcal{M}_t$ (d) $\mathcal{F}_{r,t}$ (e) $\mathcal{F} - \mathcal{F}_{r,t}$

FIG. 2.3 – Illustration du calcul d'une image-différence relative à une classe représentée par son prototype $\mathcal{M}_t$.

La ***distance de Mahalanobis*** peut aussi s'avérer utile lors de l'introduction d'espaces globaux de représentation (ACP globale ou AFD). Elle est définie comme suit :

$$\mathscr{D}_{C_t}(\mathcal{F}) = \sqrt{(\mathcal{F}_t' - \mathcal{M}_t')^T \Sigma_t^{-1} (\mathcal{F}_t' - \mathcal{M}_t')},$$

où $\mathcal{F}_t^{'}$ est le vecteur résultant de la projection de $\mathcal{F}$ sur la base W_t de la classe C_t, représentée par $\mathcal{M}_t^{'}$ et Σ_t respectivement sa moyenne et sa covariance.

Il est alors possible de définir pour l'une ou l'autre des mesures une règle de décision appropriée. A partir d'un ensemble de M classes notées $\{C_l\}_{l=1}^{M}$ et $\mathcal{F}$ un visage à identifier, nous pouvons définir pour chaque classe C_t la vraisemblance $\mathscr{L}^t = \mathscr{L}_{C_t}(\mathcal{F})$ telle que :

$$\mathscr{L}_{C_t}(\mathcal{F}) = \mathcal{N}(\mathscr{D}_{C_t}(\mathcal{F}); 0, \sigma_t),$$

où σ_t est l'écart-type des mesures de distance au sein de la classe C_t, et $\mathcal{N}(.; m, s)$ est une distribution gaussienne de moment m et de covariance s. La probabilité *a posteriori* $P(C_t|\mathcal{F}, z)$ que $\mathcal{F}$ appartienne à C_t est alors définie de la manière suivante :

$$\begin{cases} \forall t\ P(C_t|\mathcal{F}, z) = 0 \text{ et } P(C_\emptyset|\mathcal{F}, z) = 1 \text{ lorsque } \forall t\ \mathscr{L}_{C_t} < \tau \\ \forall t\ P(C_t|\mathcal{F}, z) = \frac{\mathscr{L}_{C_t}}{\sum_p \mathscr{L}_{C_p}} \text{ et } P(C_\emptyset|\mathcal{F}, z) = 0 \text{ sinon ,} \end{cases} \tag{2.5}$$

où $C_\emptyset$ représente la classe vide et τ est un paramètre libre à définir.

Comme évoqué précédemment, notre but est d'évaluer ***les SVM*** en tant que règle de décision de notre système. Considérons la forme classique du SVM [Chen et al., 2005] à deux classes labélisées $+1$ et -1. La phase d'apprentissage consiste à trouver l'hyperplan de séparation optimal (ou OSH pour *Optimal Separating Hyperplane*), séparant les deux classes dans un espace surdimensionné. Ensuite, nous pouvons définir les couples $(\mathcal{F}'^i, y_i)$, pour chaque échantillon (visage) $\mathcal{F}^i$ et $\mathcal{F}'^i$ sa projection dans le sous-espace W avec $y_i \in \{-1, +1\}$. L'équation de l'OSH est définie par $\omega * \mathcal{F}'^i + b = 0$. L'apprentissage consiste à trouver les paramètres ω et b avec $y_i(\omega.\mathcal{F}'^i + b) \leq 1$ tel que la distance entre l'hyperplan et le plus proche vecteur soit maximale.

Les données n'étant pas linéairement séparables, il est alors nécessaire de projeter les données dans un espace surdimensionné afin de pouvoir déterminer l'équation de l'OSH.

Un noyau RBF (pour *Radial Basis Function*) est généralement utilisé pour cette transformation [Heisele et al., 2001; Jonsson et al., 2000] où le paramètre γ contrôle la largeur du noyau gaussien. Un autre paramètre important à déterminer est la borne supérieure C du Lagrangien requis pour la minimisation sous contraintes. Les SVM donnent des performances très différentes suivant le choix de la fonction noyau et plus spécifiquement du choix des paramètres γ et C.

Un noyau RBF $K(\mathcal{F}'^i, \mathcal{F}'^j)$ est alors défini tel que :

$$K(\mathcal{F}'^i, \mathcal{F}'^j) = exp(-\gamma \|\mathcal{F}'^i - \mathcal{F}'^j\|^2),$$

où la marge γ est un paramètre libre.

La solution est déterminée par $\min_\alpha\{\frac{1}{2}\alpha^t Q\alpha - e^t\alpha\}$, sous les contraintes $y^t\alpha = 0$ et $0 \leq \alpha_i \leq C, i \in \{1, ..., M\}$ avec $y^t = [y_1, ..., y_M], \alpha^t = [\alpha_1, ..., \alpha_M], e^t = [1, ..., 1]$, et Q une matrice où $Q_{i,j} = y_i y_j K(\mathcal{F}'^i, \mathcal{F}'^j)$.

La règle de décision définie dans [Guo et al., 2000] est alors donnée par :

$$f(\mathcal{F}') = sign(\sum_{i=1}^{M} \alpha_i K(\mathcal{F}'^i, \mathcal{F}') + b).$$

2.1.4 Systèmes de reconnaissance et évaluations associées

Nous avons réalisé des expérimentations de reconnaissance de visages en utilisant une base de 6600 visages de 8 personnes différentes (apprises par le robot) et 3 imposteurs (non apprises). Dans ce jeu d'images de résolution fixe (30×30 pixels), les sujets bougent leur tête de manière arbitraire, changent d'expression alors que les conditions d'illumination, l'arrière plan et la distance relative changent. Un petit exemple d'images de ce jeu est montré figure 2.4.

FIG. 2.4 – Exemple d'échantillons pour une classe donnée.

De par la forte dépendance au contexte, nous n'avons pas privilégié les bases publiques de visages [Phillips et al., 2000; MIT, 2000; YALE, 2006] lors de nos évaluations. En effet, les bases publiques sont plus destinées à évaluer des applications biométriques en environnement contrôlé que des applications robotiques en environnements dynamiques. De plus, la majeure partie d'entre elles sont composées de nombreuses classes comportant un faible nombre d'images au regard de notre application. C'est pourquoi l'ensemble des évaluations seront conduites à partir de bases de visages acquises depuis notre robot mobile dans des conditions réelles (orientation naturelle du visage, expressions changeantes, changements d'illumination, etc).

Systèmes de reconnaissance évalués

1. Système FSS+EN : *Face-Specific Subspace* et norme d'erreur – Comme décrit dans [Shan et al., 2003], pour chaque classe C_t, nous calculons $W_{acp,t}$ par l'équation 2.1, et conservons $N_{v,t}$ vecteur propres (équation 2.2). La règle de décision est alors basée sur la norme d'erreur définie par l'équation 2.4. Le classifieur est entièrement défini par les ensembles $\{W^i_{acp,t}, \mathcal{M}^i_t, \sigma^i_t\}^M_{i=1}$ et paramétré par $\mathbf{q} = (\eta, \tau)'$.

2. Système GACP+MD : ACP globale et distance de Mahanalobis – Dans ce cas, une seule base de visages propres est définie suivant l'équation 2.1 et la matrice de dispersion totale S_T. La règle de décision est basée sur la distance de Mahalanobis. Ce modèle est défini par W_{acp} et les ensembles $\{\mathcal{M}'^i_t, \Sigma^i_t, \sigma^i_t\}^M_{i=1}$ et paramétré par $\mathbf{q} = (\eta, \tau)'$.

3. Système AFD+MD : Fisherface et distance de Mahanalobis – Les visages de Fisher sont déduit de l'équation 2.3 afin de définir W_{afd} comme base de projection pour y calculer la distance de Mahalanobis. Ce classifieur est entièrement défini par W_{afd} et les ensembles $\{\mathcal{M}'^i_t, \Sigma^i_t, \sigma^i_t\}^M_{i=1}$. Ce modèle dépend aussi de la définition des paramètres libres $\mathbf{q} = (\eta, \tau)'$.

4. Système GACP+SVM : ACP globale et SVM – Ce système calcule une ACP globale suivant l'équation 2.1 alors que le SVM sert de règle de décision. La théorie associée ainsi que les détails d'implémentation sont décrits dans [Wu et al., 2004]. Ce modèle est défini par W_{acp} et dépend des paramètres libres $\mathbf{q} = (\eta, C, \gamma)'$.

Protocole d'évaluation

Le protocole d'évaluation partitionne la base de données des visages en 4 jeux disjoints : (1) un jeu d'apprentissage #1 (8 utilisateurs, 30 images par classe) pour générer la base de projection (ACP ou AFD), (2) un jeu d'apprentissage #2 (8 utilisateurs, 30 images par classe) pour ca-

ractériser chaque classe (moyennes, écart-types, apprentissage du SVM), (3) un jeu d'évaluation (8 utilisateurs et 3 imposteurs, 40 images par classe) pour estimer les différents paramètres libres, (4) un jeu de test (8 utilisateurs et 3 imposteurs, 500 images par classe) afin d'évaluer les performances de chaque classifieur sur des données indépendantes.

Stratégie d'optimisation basée sur les courbes ROC

Les performances des classifieurs décrits ci-dessus sont analysées au travers de courbes ROC alors que le vecteur des paramètres $\mathbf{q}$ est sujet à l'optimisation. L'idée, avancée par Provost *et al.* dans [Provost and Fawcett, 2001], est la suivante. Il nous faut rechercher un ensemble de paramètres par le calcul de points ROC modélisant les taux de faux positifs (FPR pour *False Positive Rate*) et de vrai positifs (TPR pour *True Positive Rate*) définis par :

$$FPR = \frac{F_P}{F_P + V_N} \text{ et } TPR = \frac{V_P}{V_P + F_N},$$

où F_P, V_N, V_P et F_N représentent respectivement le nombre de faux-positifs, vrais-négatifs, vrais-positifs et faux-négatifs.

Pour un classifieur donné, l'ensemble $\mathcal{Q}$ des vecteurs des paramètres admissibles permet de générer un ensemble de points ROC, dans lequel nous pouvons extraire les plus dominants en terme de Pareto optimalité.

Ainsi, nous recherchons le sous-ensemble $\mathcal{Q}^* \subset \mathcal{Q}$ des vecteurs $\mathbf{q}$ pour lesquels il n'existe pas d'autres vecteurs qui surclassent chaque objectif $\mathcal{O} = \{FRR, FPR\}$:

$$\mathcal{Q}^* = \{\mathbf{q} \in \mathcal{Q} | \forall\, \mathbf{q}^{'} \in \mathcal{Q}, \forall f_1 \in \mathcal{O}, f_1(\mathbf{q}) \geq f_1(\mathbf{q}^{'}) \wedge\ \exists f_2 \in \mathcal{O}, f_2(\mathbf{q}) > f_2(\mathbf{q}^{'})\}, \qquad (2.6)$$

où FRR (pour *False Rejection Rate*) correspond au taux de faux-négatifs tel que $FRR = 1 - TPR$.

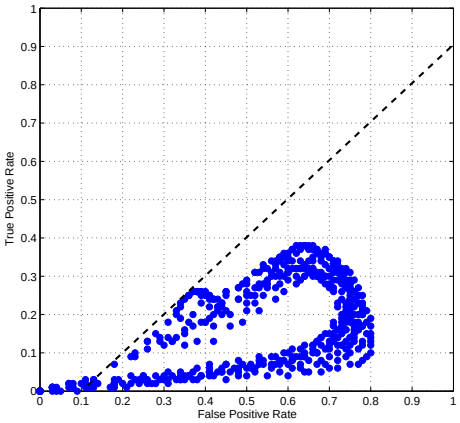
Plus clairement, $\mathcal{Q}^*$ identifie le sous-ensemble des vecteurs qui sont potentiellement optimaux pour un classifieur donné.

Généralement, un classifieur est défini pour une sensibilité (ou TPR) donnée. Les points ROC peuvent alors être utilisés pour choisir les meilleurs points d'opération, *i.e.* correspondant aux performances désirées. Ce point idéal doit être choisi de manière à offrir le meilleur compromis entre le taux de faux négatifs et le taux de faux positifs. Le coût moyen de classification pour un point (x, y), d'abscisse FPR et d'ordonnée TPR, de l'espace ROC est $C = (1-p)\alpha x + p\beta(1-y)$ où p est un facteur de proportion entre faux-positifs et faux-négatifs, et α et β sont respectivement les coûts d'obtenir un faux-positif et un faux-négatif [Provost and Fawcett, 2001]. Le gradient de chaque ligne d'isocoût dépend des rapports $\frac{\alpha}{\beta}$ et $\frac{(1-p)}{p}$. Si les facteurs de coût sont égaux, *i.e.* $\alpha = \beta$ et la proportion entre faux-positifs et faux-négatifs égale 50%, *i.e.* $p = 0.5$, le gradient est égal à 1 et la ligne d'isocoût est à $45°$.

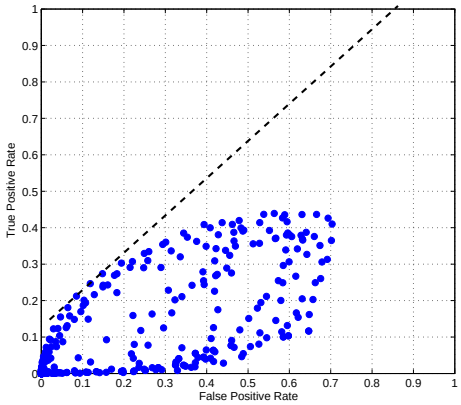
L'EER (pour *Equal Error Rate*) est alors déduit de la ligne d'iso-coût où C représente la moitié des faux-positifs plus la moitié des faux-négatifs tel que :

$$EER = 0.5x + 0.5(1 - y).$$

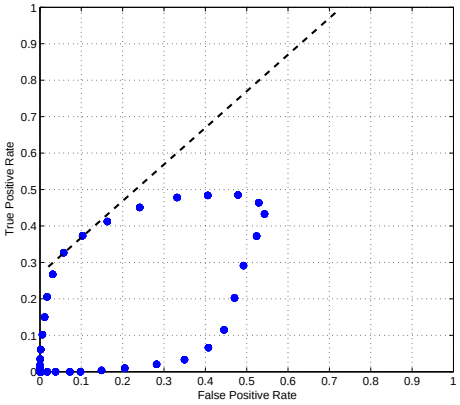
En considérant un nombre égal de positifs et de négatifs pour un classifieur donné, ce taux correspond donc au point de l'espace ROC situé sur la ligne d'isocoût à $45°$ la plus proche du coin haut-gauche.



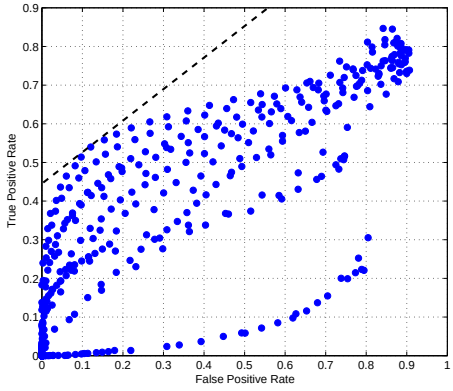
(a) Système FSS+EN, $\mathbf{q} = (\eta, \tau)'$: EER=0.51.



(b) Système GACP+MD, $\mathbf{q} = (\eta, \tau)'$: EER=0.44.



(c) Système AFD+MD, $\mathbf{q} = (\eta, \tau)'$: EER=0.37.



(d) Système GACP+SVM, $\mathbf{q} = (\eta, C, \gamma)'$: EER=0.29.

FIG. 2.5 – Points ROC pour chaque classifieur et la ligne d'iso-coût associée à l'EER. Le vecteur de paramètres $\mathbf{q}$ à optimiser est listé en dessous de chaque classifieur.

Les performances des classifieurs décrits figure 2.5 sont similaires en termes de spécificité (ou TNR pour *True Negative Rate*) *i.e.* :

$$TNR = \frac{V_N}{F_P + V_N} \sim 90\%.$$

Ces résultats sont très prometteurs au sens où tous les classifieurs sont robustes aux variations de poses, illumination et expression des visages testés. A contrario, il existe des différences

significatives entre les systèmes observés en terme de faux-positifs (ou FPR) et faux-négatifs (ou FRR).

La figure 2.5 montre les points ROC ainsi que le front de Pareto issus de la variation des paramètres. Chaque sous-figure permet d'observer et de comparer les performances relatives des classifieurs.

Au regard du FPR, le système AFD+MD (figure 2.5(c)) présente un maximum de ~ 0.55 alors que les autres systèmes atteignent jusqu'à ~ 0.9. Ce système permet d'obtenir une classification très fiable mais impliquant de nombreux rejets ($FPR = 0.02\ vs.\ TPR = 0.3$). De plus, l'utilisation d'une AFD permet une certaine flexibilité quand au choix du paramètre η. En effet, l'AFD visant à séparer au mieux l'ensemble des classes, la majorité de *l'information* importante est contenue dans les premiers vecteurs propres *i.e.* pour un η petit. L'augmentation de η ne fait donc qu'ajouter une infime partie d'information, pas assez importante pour faire varier les performances de classification.

Globalement, le système GACP+SVM domine clairement les autres classifieurs puisque son front de Pareto est le plus proche du coin haut-gauche de l'espace ROC (correspondant au classifieur idéal).

Dans ce cas, le meilleur système, le système GACP+SVM, offre un front de Pareto avec la plus basse EER, soit 0.29. Il faut noter que ce système peut traiter 2000 images par seconde contre un peu plus de 3000 images par seconde pour les autres systèmes. En effet, la classification par SVM est plus coûteuse en temps de calcul par rapport à une classification utilisant la norme d'erreur ou la distance de Mahalanobis. Cette faiblesse est tout de même compensée par une faible complexité de son algorithme de classification. Là où les algorithmes présentés figures 2.5(a-c) ont une complexité grandissante en $\mathbb{O}(n)$ (n étant le nombre de classes apprises), l'algorithme basé sur une classification par SVM (figure 2.5(d)) reste de complexité constante *i.e.* qui ne dépend pas du nombre de classes mises en jeu.

Malheureusement, une recherche exhaustive par analyse de courbes ROC pour choisir les meilleurs paramètres n'est pas viable pour un robot autonome vu que la finalité est de pouvoir apprendre un visage humain en ligne, lors de la première interaction entre le robot et l'homme. Par conséquent, nous proposons d'utiliser une approche par algorithme génétique afin de trouver le vecteur de paramètres $\mathbf{q}$ optimal pour le système GACP+SVM de manière plus efficace et rapide grâce à l'optimisation multi-objectifs. En limitant le nombre de points ROC considérés, les algorithmes génétiques rendent la procédure d'optimisation compatible en temps de calcul par rapport au contexte applicatif.

Stratégie d'optimisation basée sur un algorithme génétique

Les méthodes conventionnelles utilisant les algorithmes génétiques traitent du problème d'optimisation mono-critère [Seo, 2007]. L'algorithme génétique NSGA-II (pour *Non-dominated Sorting GA*) est mieux adapté aux problèmes d'optimisation multi-critères [Xu and Li, 2006] puisqu'aucune solution ne peut atteindre un optimum global pour plusieurs objectifs, *i.e.* minimiser le taux de faux positif (ou FPR) et le taux de faux négatifs (ou FRR). Si l'évaluation du

premier objectif ne peut être améliorée sans dégrader l'évaluation du second, la solution globale se rapporte à l'ensemble des solutions Pareto-optimales ou non-dominées [Gavrila and Munder, 2007].

L'algorithme 2.1 décrit les étapes du processus utilisé pour optimiser les paramètres libres. Après une phase d'initialisation visant à générer une population $P_0 = \{\mathbf{q}_0^i\}_{i=1}^N$ de N vecteurs de paramètres (ou individus) (étape 1), une population naissante est créée à chaque itération t par des méthodes de croisement et de mutation (étape 3). Un croisement consiste à sélectionner aléatoirement certains paramétres d'un individu $\mathbf{q}_{t-1}^a$ et de le compléter avec ceux d'un individu $\mathbf{q}_{t-1}^b$ afin de créer deux nouveaux individus $\mathbf{q}_t^{a'}$ et $\mathbf{q}_t^{b'}$. Une mutation modifie aléatoirement un seul paramètre d'un individu $\mathbf{q}_{t-1}^a$ pour donner un nouveau $\mathbf{q}_t^{a'}$. Une fois cette population générée, l'ensemble des solutions Pareto-optimales, tel que défini par l'équation 2.6, est identifié par l'ensemble des individus non-dominés (étape 5). Afin d'éviter le confinement des solutions sur un optimum local, une étape de "désengorgement" est appliqué en ne sélectionnant que les individus du font de Pareto assez distants les uns des autres (étapes 6 à 13). Cette approche nous permet de trouver le meilleur compromis entre FPR et FRR en trouvant le front de Pareto

$$F_i = \{f(\mathbf{q}) \in \mathcal{O} | \mathbf{q} \in \mathcal{Q}^*\},$$

avec $\mathcal{Q}^*$ l'ensemble des solutions Pareto-optimales.

ALG. **2.1** Algorithme NSGA-II.

1: Créer une population parent P_0 de taille N. $t \leftarrow 0$
2: **répéter**
3: Appliquer les croisements et mutations à P_t pour créer une population naissante $\mathcal{Q}_t$ de taille N.
4: $R_t = P_t \cup \mathcal{Q}_t$.
5: Identifier le front non-dominé $F_1, F_2, \ldots, F_k$ dans R_t.
6: **pour** $i = 1, \ldots, k$ **faire**
7: Calculer la distance d'encombrement des solutions de F_i. Trier par encombrement.
8: **si** $|P_{t+1}| + |F_i| > N$ **alors**
9: Ajouter les $N - |P_{t+1}|$ solutions les moins encombrées à P_{t+1}.
10: **sinon**
11: $P_{t+1} = P_{t+1} \cup F_i$.
12: **fin si**
13: **fin pour**
14: Sélectionner *par tournois* basé sur la distance d'encombrement les parents depuis P_{t+1}.
15: **jusqu'à** Equation 2.6 non satisfaite

L'algorithme a pour but de minimiser la distance entre les solutions générées et le front de Pareto (2.6) et de maximiser la diversité de l'approximation du front de Pareto. La figure 2.6 montre l'évolution du front de Pareto suivant un ensemble de vecteurs de paramètres tels que $card(\mathcal{Q}) \in [16, 20]$ et le nombre de génération t est inférieur à 30.

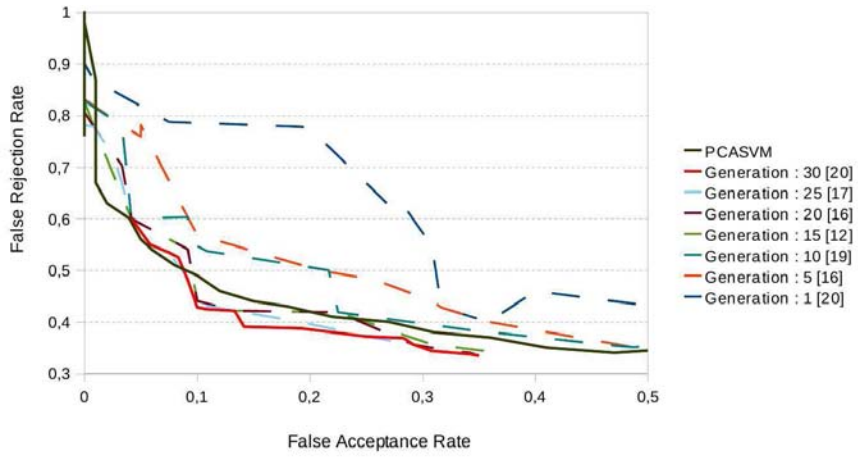


FIG. 2.6 – Evolution du front de Pareto NSGA-II relatif au système GACP+SVM.

On remarque que les solutions générées évoluent dans l'espace ROC afin de réduire les deux objectifs (FPR et FRR). Cette stratégie d'optimisation ne garantit pas de trouver le front de Pareto optimal, mais il est évident que la solution empirique est proche d'un optimum ne dépendant que du nombre de générations. A partir d'une population initialisée aléatoirement (génération n°1 sur la figure 2.6), nous pouvons voir qu'après les dix premières générations, il existe déja une solution qui surpasse celles sans optimisation alors que les 30 générations suivantes n'améliorent que faiblement ce précédent résultat. De plus, l'EER minimum aprés 30 générations est de 0.26 contre 0.29 dans la figure 2.5(d). Pour information, cette stratégie de recherche non exhaustive, paramétrée par la taille des générations et par le nombre de ces dernières, permet de définir le compromis entre le coût, le temps de calcul et les performances de classification.

Afin d'évaluer le gain de performance, le jeu de test, comprenant 8 utilisateurs et 3 imposteurs avec 500 images par classe, a été soumis à classification avec les paramètres libres issus des deux types d'optimisation. L'utilisation des paramètres libres, optimisés par l'algorithme génétique et listés en table 2.1, a permis de réduire le taux de faux-positifs de 0.12 à 0.06 alors que le taux de vrai-positifs ne décroit que de 5% (0.54 *vs.* 0.49).

Symbole	Description	Valeur
η	Energie des vecteurs propres de l'ACP	0.99
C	Borne supérieure du Lagrangien du SVM	80391
γ	Largeur du noyau RBF du SVM	0.002526

TAB. 2.1 – Valeur des paramètres utilisés pour le système GACP+SVM.

2.2 Identification de personnes par radio fréquence

La majorité des systèmes d'identification visuelle, comme les méthodes de détection / reconnaissance de visages [Germa et al., 2007b; Viola and Jones, 2003], font l'hypothèse que la personne regarde le robot. Leurs performances dépendent énormément des conditions d'illumination, de l'angle de vue, de la distance à la cible ainsi que de la variabilité de l'apparence humaine dans une vidéo. Par conséquent, de nombreuses approches considèrent des capteurs non nécessairement visuels. Leur but est de combiner les informations issues de différents capteurs avec des données issues d'un flux vidéo.

2.2.1 Considérations générales

Certaines approches se focalisent sur les "technologies émergentes" basées sur les réseaux sans fil et les ultrasons, les infrarouges [Schulz et al., 2003] ou les badges Radio Fréquence (RF) [Kanda et al., 2007]. Au contraire des premières techniques citées ne permettant que de détecter les personnes sans pouvoir les identifier, les badges RF sont équipés d'une électronique dédiée capable de capter la puissance des ondes radio émises depuis un lecteur à une certaine distance. Cette puissance leur permet ensuite d'envoyer une réponse contenant un identifiant unique capté par le lecteur au travers d'une antenne.

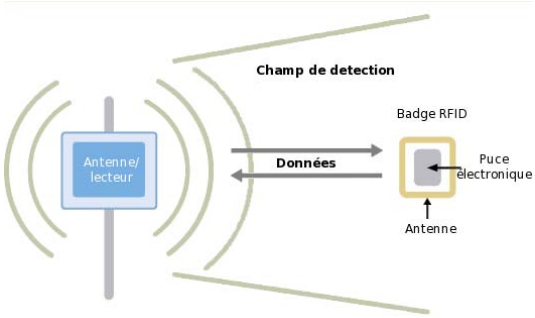


FIG. 2.7 – Schéma de principe de la RFID.

L'identification par radio-fréquence (ou RFID pour *Radio Frequency IDentification*) est une technologie d'identification parmi un large éventail de technologies d'identification telles que les codes à barres ou les technologies d'identification biométriques.

Le principe est assez simple. Les RFID utilisent les ondes radios pour recevoir et transmettre des données stockées dans la puce, présente dans le badge. Il existent de nombreuses applications industrielles qui utilisent la technologie RFID : le suivi de production, les systèmes de paiement, les contrôles d'accès sécurisés, etc. L'information peut être stockée dans un badge. Une fois le badge activé par les ondes radios émises par une antenne depuis un lecteur, l'information est retournée à l'antenne au travers de la même onde porteuse (figure 2.7). Un badge RF est constitué d'une puce de silicium et d'une antenne. La puce contient généralement un identifiant unique et éventuellement d'autres informations dépendant des propriétés de stockage de la puce. Le type d'application permet de déterminer le type de badges à utiliser parmi les badges passifs, semi-passifs et actifs.

Les badges passifs ne possèdent pas de batterie. Lorsque le badge reçoit un stimuli provenant d'un lecteur, il emmagasine de l'énergie au travers de son antenne. Cette énergie est ensuite utilisée pour renvoyer l'information contenue dans le badge. L'absence de source d'alimentation propre ne permet pas aux badges passifs d'atteindre des portées trop élevées. En général, ces

dernières se situent entre quelques centimètres comme dans les passeports et quelques mètres (clé électronique pour l'automobile). Les badges actifs sont eux beaucoup plus élaborés, mais aussi plus onéreux, plus fragiles avec une durée de vie moindre que leurs équivalents passifs ayant une durée de vie illimitée. Ils possèdent une batterie interne leur permettant de renvoyer une information sur des distances plus importantes pouvant aller jusqu'à plusieurs centaines de mètres

De tels badges peuvent alors délivrer une information explicite sur l'identité d'une personne même si l'information concernant la position est très vague.

2.2.2 Etat de l'art

Depuis plusieurs années, les RFID connaissent une utilisation grandissante dans de nombreux domaines notamment en robotique. Dans [Kubitz et al., 1997], Kubitz *et al.* utilisent les badges RFID comme point de repère d'une carte topologique. Plus récemment, Kulyukin *et al.* [Kulyukin et al., 2004] proposent un système robotique basé sur les RFID pour l'assistance aux malvoyants. Des badges passifs sont alors associés à des objets de l'environnement et permettent d'étudier les comportements locaux. Bien qu'efficaces pour la navigation de robots mobiles, les approches RFID proposées dans la littérature supposent que la position du badge est connue *a priori*. Cette hypothèse, bien que suffisante dans des environnements statiques et connus tel que les applications industrielles, n'est pas concevable dans des environnements humains, comme les musées ou les supermarchés, ou pour localiser aussi bien les objets du quotidien que les personnes équipées de badges. Pour palier à ce problème, les badges RF actifs peuvent alors être utilisés [Chae and Han, 2005; Zhou et al., 2007]. Cependant, nous avons vu que leurs caractéristiques ne sont pas adaptées à notre contexte applicatif.

Il existe de nombreuses stratégies de localisation de badges passifs dans le contexte des robots mobiles. Hähnel *et al.* [Hahnel et al., 2004] suggèrent d'utiliser une approche Bayesienne afin d'estimer la position d'un badge passif depuis un robot mobile équipé d'un lecteur RFID et d'un laser. Les badges sont aussi utilisés pour améliorer les performances de localisation par des méthodes de Monte Carlo. Une approche Bayesienne similaire est utilisée par Liu *et al.* [Liu et al., 2006] afin de localiser des objets mobiles. Cette méthode est implémentée sur une plateforme mobile équipée d'un capteur RFID et d'une caméra. Enfin, Jia *et al.* [Jia et al., 2006] utilisent les badges RFID afin de détecter et d'éviter les obstacles. La règle Bayesienne est appliquée pour estimer la position du badge. Ces derniers sont ensuite utilisés comme points de repère pour la localisation du robot basée sur des indices visuels issus d'une tête stéréo. Bien qu'offrant des performances intéressantes pour la localisation de cibles fixes, ces systèmes RFID ne prennent pas en compte un déplacement de la cible et ne sont donc pas adaptés à notre utilisation.

Les applications visant à utiliser les RFID pour percevoir les activités humaines impliquent principalement des capteurs répartis dans l'environnement, aussi appelés capteurs ubiquistes [Koch et al., 2007; Chen et al., 2009; Lieckfeldt et al., 2009]. Koch *et al.* dans [Koch et al., 2007] proposent de localiser humains, objets et robots mobiles au sein d'un environnement vivant équipé de badges actifs et passifs afin d'induire un comportement intelligent de l'environ-

nement. De plus, les techniques de traitements du signal sont souvent employées pour palier au manque de précision de la technologie RFID. L'une des plus utilisées est la méthode mesurant la force du signal reçu (ou RSS pour *Recieved Signal Strength*) [Lieckfeldt et al., 2009]. Ni *et al.* utilisent un système RFID actif afin de localiser une cible mobile équipée d'un badge RFID [Ni et al., 2004]. Les lecteurs RFID stationnaires déployés dans l'environnement comparent le niveau de puissance mesuré sur des badges de référence afin d'améliorer les performances de localisation. Un autre système de localisation basé sur les RFID est SpotOn qui permet de définir et construire un matériel dédié pour la localisation par RFID. Un algorithme de localisation 3D utilise la RSS pour déterminer la position de chaque badge. En ce qui concerne les capteurs ubiquistes, l'utilisation de ressources perceptuelles déportées permet en général d'améliorer la performance globale du système. Cependant, notre capteur devant être embarqué sur le robot afin de réduire l'équipement de l'environnement, il est difficile de concevoir de tels systèmes.

Jia *et al.* [Jia et al., 2008] proposent d'utiliser un capteur RFID et un banc de stéréovision embarqués sur un robot mobile afin de localiser les personnes présentes dans l'environnement. Le système RFID multi-antennes permet de définir une zone d'observation alors que la vision permet d'afiner les mesures effectuées sur l'homme. Ferret considère le problème de la localisation d'objets en mouvement et utilise l'aspect directionnel du lecteur RFID [Liu et al., 2006]. L'idée est d'exploiter différentes orientations du lecteur pour afiner l'estimation de la position d'un badge. Cette approche utilise, elle, un système RFID passif. A notre connaissance, très peu d'applications considèrent un capteur RF embarqué pour détecter les personnes. Dans ce cas, la zone de détection est limitée à $180°$ et un seul capteur est utilisé : le badge RFID.

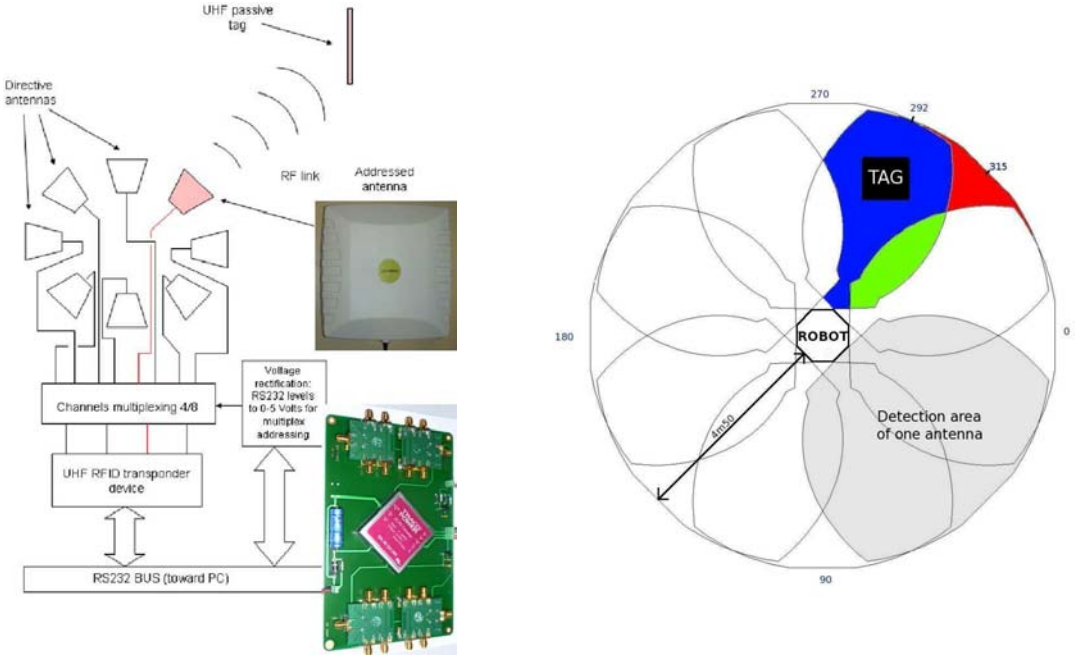
Une dernière remarque concerne la mise en place du capteur RFID. Notre application privilégie des ressources perceptuelles embarquées dans le but de limiter le coût d'installation de matériel dans l'environnement et leur entretien. Dans ce but, nous avons personnalisé un système RFID du commerce, pour s'appuyer sur l'existant, afin de le rendre capable de détecter des badges RF sur $360°$ grâce au multiplexage de 8 antennes. L'avantage est de pouvoir détecter et identifier les personnes au voisinage du robot et indépendamment de leur situation par rapport au robot. Nous avons ensuite embarqué ce système sur notre robot mobile Rackham afin de détecter les personnes équipées de badges passifs. Ce capteur omnidirectionnel peut alors être considéré comme un complément idéal à notre système de vision monoculaire possédant, par définition, un champ de vue limité.

2.2.3 Description de notre système

Description matérielle

Un système expérimental RFID a été développé afin de réaliser une étude de faisabilité dans notre contexte applicatif. Le capteur se compose de : (i) un lecteur RFID multi-protocôles du marché (CAENRFID $A941$) fonctionnant à 870MHz, (ii) 8 antennes RFID directionelles capables de détecter un badge RF passif porté par l'utilisateur, (iii) un prototype de carte de multiplexage RF afin d'adresser les 8 antennes de manière séquentielle (figure 2.8(a)). En effet, à partir d'une seule antenne, seul l'angle d'ouverture de l'antenne permet d'estimer la position

d'un badge. Avec plusieurs antennes, le badge RF peut être détecté tout autour du robot. Connaissant la position des antennes ainsi que leur angle d'ouverture, la zone autour du robot peut être découpée en différentes zones, suivant le nombre d'antennes détectant simultanément un badge.



(a) (b)

FIG. 2.8 – Prototype de multiplexage RF pour 8 antennes (a) et les zones de détection associées (b). La zone grisée représente la zone de détection d'une antenne.

Le lecteur RFID $A941$ de CAENRFID est un module de lecture multi-protocôle longue portée développé pour des applications intégrées à une infrastucture utilisateur impliquant la lecture et la programmation de bagde Ultra Haute Fréquence (ou UHF). Le lecteur $A941$ est disponible en différentes version. La version ETSI est entièrement conforme aux régulations européennes (ETSI EN 302 208 et ETSI EN 300 220). Le lecteur est disponible en tant que module OEM, pour les utilisateurs qui souhaitent développer leur propre solution RFID et requièrent un lecteur longue portée, ou en module IP65 permettant une meilleure embarquabilité de l'ensemble. Ce lecteur est équipé de quatre ports RF permettant de connecter quatre antennes ainsi que d'un port RS232 permettant de dialoguer avec un système extérieur.

Les antennes RFID sont des antennes à polarisation linéaire couvrant la gamme de fréquence comprise dans $[860-970]$MHz. Leur ouverture est de $69°$ sur le plan horizontal et $65°$ sur le plan vertical. Leur dimension est de 22×22cm^2. La polarisation linéaire nécessite une orientation donnée du badge (grossièrement verticale) pour qu'il soit détecté mais permet d'augmenter la portée des antennes par rapport à une antenne à polarisation circulaire.

Le duplexeur RFID permet de dupliquer le nombre d'entrées du lecteur RFID. Il est alors possible d'utiliser 8 antennes au lieu de 4 prévues initialement (figure 2.8(a)). Le but de ce duplexeur est de permettre au lecteur d'être connecté de façon intermittente à l'une ou l'autre moitié des antennes. Il est composé de 4 commutateurs RF (figure 2.9) nécessitant une tension de référence ainsi qu'une commande de commutation reliée à une broche libre du connecteur RS-232. Ce duplexeur a été réalisé au laboratoire par l'équipe 2I en collaboration avec l'équipe MINC.

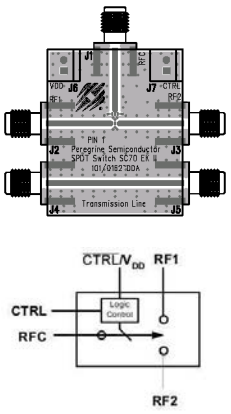


FIG. 2.9 – Image (a) et schéma (b) d'un commutateur RFID.

Modèle de mesure du capteur

Ce système nous permet de discrétiser l'espace autour du robot en 24 zones dépendant uniquement du nombre d'antennes qui détectent simultanément le même badge (figure 2.8(b)). Un même badge peut donc être détecté tout autour du robot dans un périmètre allant de 0.5m (*i.e.* diamètre approximatif du robot) à 4.5m (*i.e.* portée maximale de l'antenne), ce qui correspond aux distances sociales usuelles mises en jeu lors d'interactions Homme / Robot. Afin de caractériser le modèle d'observation de la ceinture d'antennes, nous avons réalisé une étude dépendant du nombre d'antennes détectant un même badge. Les histogrammes normalisés associés sont présentés sur la figure 2.10 sachant que l'axe x représente respectivement l'azimut (noté θ dans (a), (b), (c)) et la distance (notée ρ dans (d), (e), (f)).

Le modèle de capteur résultant de cette étude nous permet d'approximer les histogrammes en distance et en azimut par un modèle gaussien respectivement annoté $(\mu_\theta^{tag}, \sigma_\theta^{tag})$ et $(\mu_\rho^{tag}, \sigma_\rho^{tag})$ où $\mu_{(.)}^{tag}$ et $\sigma_{(.)}^{tag}$ correspondent à la moyenne et à l'écart-type de chaque mesure. Il est donc possible de calculer la probabilité de chaque position (ρ, θ) du badge RF sur la carte de saillance (figure 2.11). A partir de ce modèle, nous avons caractérisé les performances globales du capteur.

2.2.4 Mise en œuvre et évaluations associées

Le système RFID a été monté sur la plateforme mobile Rackham (présentée chapitre 5) et évalué dans un contexte encombré. Nous avons alors évalué le nombre de vrais positifs (détections effectives) dans une zone de 81m^2 autour du robot. Des obstacles ont été ajoutés aléatoirement selon une loi uniforme dans la zone de détection des antennes de façon incrémentale. La vérité terrain est basée sur le rapport entre les zones occultées par les obstacles et la surface totale de la zone. Nous avons ensuite reproduit cette situation sur le terrain en observant les résultats de détection d'un badge en fonction du nombre d'obstacles.

La comparaison entre les données expérimentales et théoriques est présentée figure 2.12.

Les axes x et y de la figure 2.12 représentent respectivement le nombre de personnes susceptibles d'occulter le badge (*i.e.* l'encombrement de la scène) et le taux de détection. Les rectangles ainsi que les lignes centrales indiquent le degré de dispersion (pour 50% des mesures) et la médiane. Nos courbes expérimentales sont assez proches des résultats théoriques. Le système

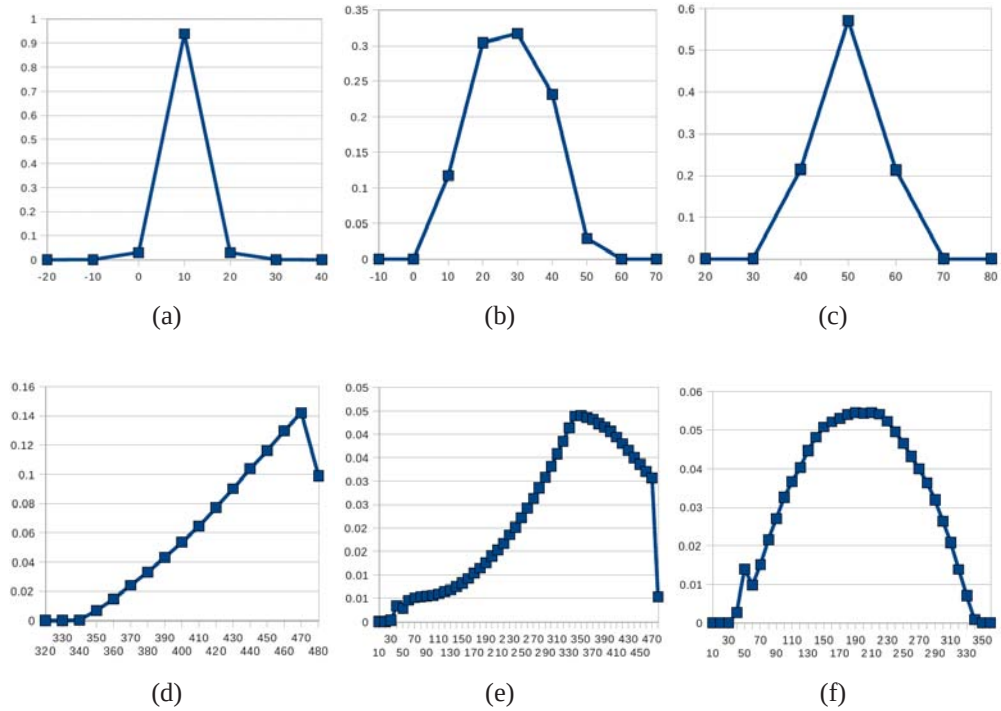


FIG. 2.10 – Histogrammes d'occurrence de détection en azimut (a, b, c) et distance (d, e, f) pour la détection d'un badge par une (a-d), deux (b-e) ou trois (c-f) antennes.

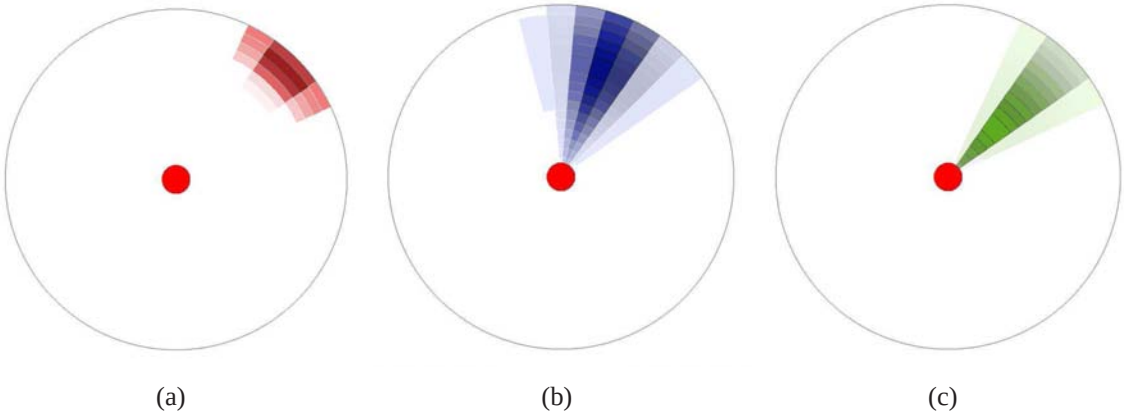


FIG. 2.11 – Cartes de saillance pour la détection d'un badge par 1 (a), 2 (b) ou 3 (c) antennes.

étant perturbé par les occultations, le nombre de faux-négatifs croît logiquement avec le nombre d'obstacles. Malgré tout, le taux de détection demeure satisfaisant, même pour des scènes plutôt encombrées (*e.g.* 70% de détection en moyenne pour 7 personnes présentes autour du robot).

De plus, très peu de faux-positifs sont observés en pratique du fait que le badges passifs induisent moins de réflections dans l'environnement que leur version active.

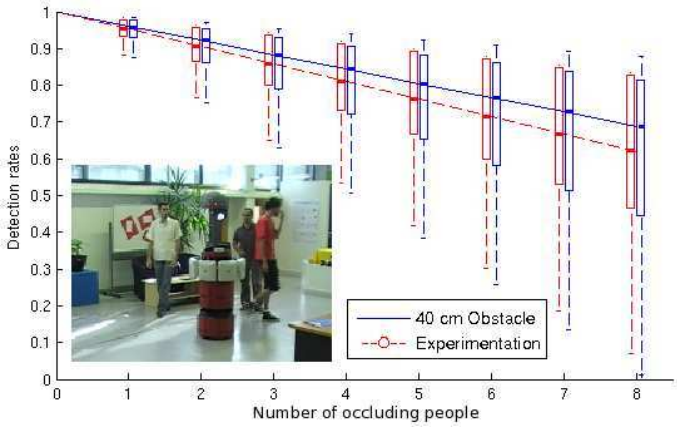


FIG. 2.12 – Taux de détection *vs.* encombrement autour du robot.

2.2.5 Vers un capteur plus compact

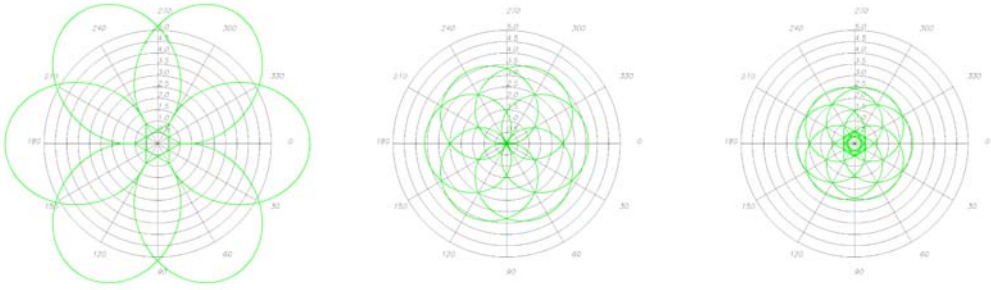
Etude théorique

Dans le but d'améliorer la compacité et l'embarquabilité de notre capteur RFID omnidirectionnel, il nous faut dimensionner un nouveau système RFID. D'après la forme hexagonale de notre plateforme Rackham décrite chapitre 5, nous avons étudié l'intégration de six antennes. Ceci nous permet de réduire le volume global du système afin de réduire son encombrement sur le robot tout en conservant une couverture maximale autour de ce dernier.

La disposition circulaire des antennes nous permet alors d'obtenir plusieurs zones de recouvrement, chacune dépendant du nombre d'antennes pouvant détecter un badge situé dans cette dernière. Il est alors nécessaire de dimensionner la zone de détection d'une antenne afin d'optimiser le nombre de zones de recouvrement mais aussi d'uniformiser leur taille relative. Il faut alors noter que la définition d'un système idéal est un compromis entre l'angle d'ouverture des antennes et la distance de détection permettant d'identifier un badge dans la plage de distance définie pour une interaction sociale Homme / Robot.

Différentes simulations ont alors été réalisées afin d'obtenir un système idéalement dimensionné pour notre application. La figure 2.13 présente les résultats de simulation pour différents angles d'ouverture d'une antenne RFID.

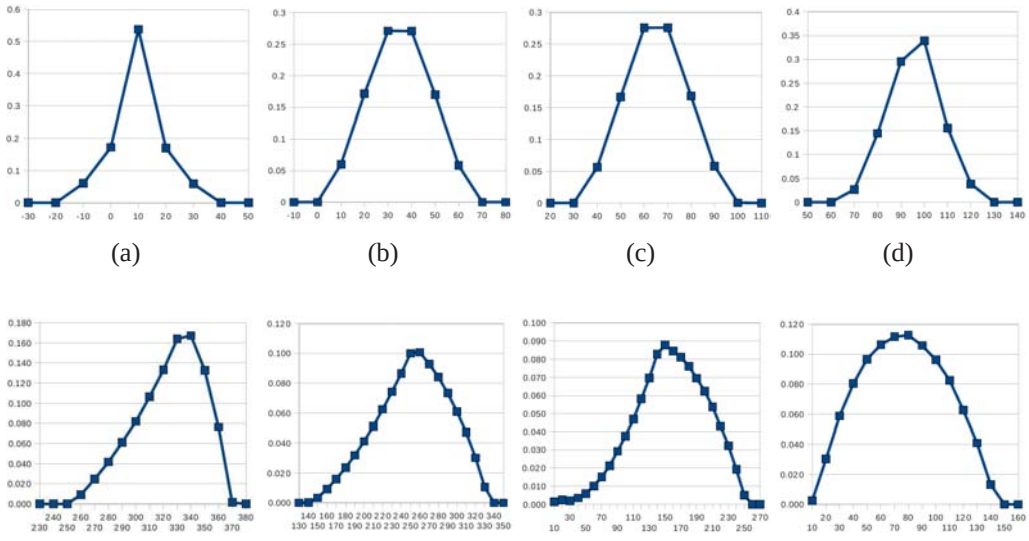
Le modèle présenté figure 2.13(b), relatif aux antennes ayant un angle d'ouverture de $180°$, est privilégié car il est capable de détecter un badge à une distance comprise dans $[0; 3.5]$m, correspondant à une distance sociale entre le robot et le badge. De plus, les zones de recouvrement



(a) (b) (c)

FIG. 2.13 – Simulation de l'ensemble des zones de recouvrement pour des antennes avec un angle d'ouverture de $90°$ (a), $180°$ (b) et $270°$ (c).

sont assez nombreuses et étroites pour localiser le badge plus précisément que pour le système à huit antennes.



(a) (b) (c) (d)

(e) (f) (g) (h)

FIG. 2.14 – Histogrammes d'occurrence de détection en azimut (a, b, c, d) et distance (e, f, g, h) pour la détection d'un badge par une (a-e), deux (b-f), trois (c-g) ou quatre (d-h) antennes.

Les histogrammes normalisés associés à ce modèle de capteur, présentés figure 2.14, montrent bien une certaine homogénéité dans la taille des zones de recouvrement des antennes, tant en distance ρ qu'en azimut θ.

2.3 Vers l'intégration de détecteurs complémentaires

Les systèmes de détection et d'identification de personnes ne peuvent se résumer aux deux seules approches présentées. En effet, l'identification faciale de personnes reste dédiée à une interaction proche (2.5m maximum) alors que le système RFID fourni une information vague de la position de l'utilisateur. De plus, la tâche d'évitement, et donc de suivi multi-cibles, requiert d'autres détecteurs de plus grande portée. Nous avons développé deux détecteurs supplémentaires qui nous semblent complémentaires avec ceux présentés plus haut, à savoir (1) un détecteur basé sur des données issues d'un laser [João et al., 2005], (2) un détecteur visuel de personnes basé sur un classifieur SVM de silhouette récemment introduit dans [Felzenszwalb et al., 2009]. Le premier utilise un capteur très répandu pour la navigation métrique en robotique alors que le dernier se justifie par le fait que les passants ne font pas forcément face au robot et il n'est donc pas possible de détecter leur visage. Les travaux présentés dans cette section ont été l'objet du stage de Master de Alhayat Ali Mekonnen.

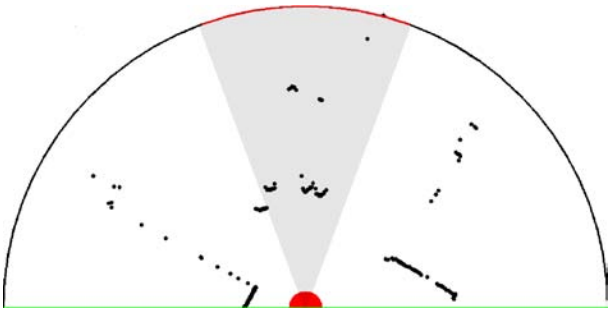
2.3.1 Detection laser

La détection de personne par laser a été largement abordé ces dernières années dans la littérature. L'idée principale est de détecter les jambes d'une (ou plusieurs) personne située autour du robot au moyen d'un laser. En effet, un laser mesure la distance radiale d'un obstacle situé dans un arc de 180° avec une précision de 0.5°. Par conséquent, une personne située dans le champ de vue d'un laser correspond à un contour spécifique relatif à ses jambes. Il est alors possible de déterminer la présence et la position d'une personne par la détection de ses jambes. Une illustration du processus de détection est présentée figure 2.15.

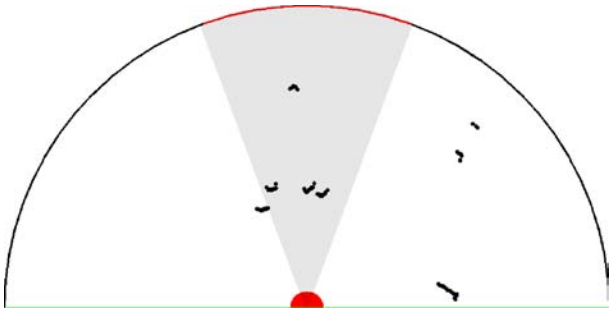
Afin de segmenter une jambe à partir de données laser (figure 2.15(b)), des contraintes géométriques correspondant au contour d'une jambe sont utilisés [João et al., 2005]. Dans un premier temps, les points candidats sont regroupés selon un critère de distance entre deux points consécutifs. Une étape de filtrage est alors introduite afin d'éliminer les groupes contenant (i) des points alignés, (ii) un nombre de points ne correspondant pas à un contour de jambes (figure 2.15(c)). Par la suite, pour chaque groupe, les angles inscrits formés à partir des deux points extrêmes et de chaque point intermédiaire sont calculés. La moyenne et l'écart-type de ces angles internes permet alors de caractériser l'arc de cercle défini par ces points. Comme décrit dans [João et al., 2005], les contours d'une jambe correspondent à une moyenne comprise entre 90° et 135° et un écart-type de 8.5°. Les groupes ne correspondant pas à ces spécifications sont alors éliminés. L'étape suivant consiste à regrouper les groupes candidats par paire lorsque ces derniers sont suffisamment proche l'un de l'autre. Chaque paire correspond alors à une personne positionnée au centre de chaque paire (figure 2.15(d)).

Le système permet de définir, pour chaque personne $\mathcal{P}$, sa position $(\mu_x^{\mathcal{P}}, \mu_y^{\mathcal{P}})$ dans le repère du robot ainsi que son écart-type $(\sigma_x^{\mathcal{P}}, \sigma_y^{\mathcal{P}})$ relatif à la précision du capteur. Il est donc possible de calculer la probabilité de chaque position (x_i, y_i) sur la carte de saillance (figure 2.15(e)) par une mixture de gaussienne comme précédemment.

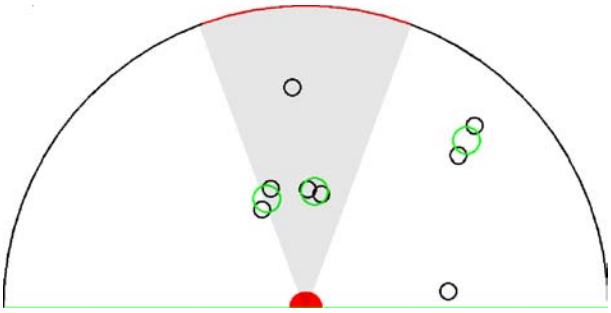
(a) Situation Homme / Robot.



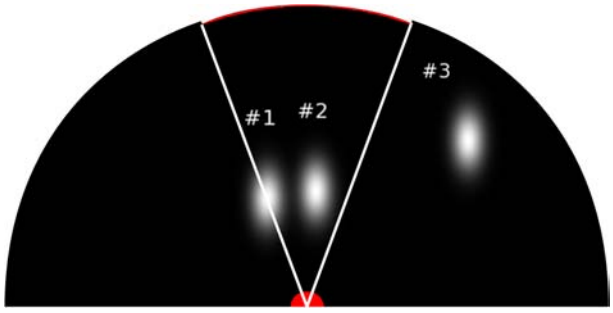
(b) Données brutes issues du laser.



(c) Données candidates après filtrage.



(d) Position des jambes (cercles blancs) et des personnes (cercles verts) segmentées.



(e) Carte de saillance associée aux détection laser.

FIG. 2.15 – Illustration du processus de segmentation des jambes à partir de données laser. Le disque rouge correspond au robot alors que le cercle blanc délimite la zone de détection du laser (4.5m) et l'arc de cercle rouge correspond au champ de vue de la caméra.

2.3.2 Détection visuelle de personnes

Cette sous-section présente succintement un détecteur de personnes développé par Felzenszwalb *et al.* [Felzenszwalb et al., 2009] très adapté à notre contexte car très robuste aux occultations. Ce détecteur est basé sur un mélange de modèles de parties déformables ayant la par-

ticularité de représenter des classes d'objets hautement variables *e.g.* les différentes parties du corps. Le détecteur résultant donne de bons résultats (en rapport avec l'état de l'art) sur les bases d'images du challenge PASCAL VOC [Everingham et al., a; Everingham et al., b] en terme d'efficacité et de précision.

Dans un premier temps, il est nécessaire de définir un modèle correspondant à une personne afin d'en spécifier les composantes individuelles à extraire. Ensuite, une phase d'apprentissage est indispensable avant d'utiliser le modèle appris pour détecter une personne.

Un exemple des étapes de détection est présenté figure 2.16. Le modèle que nous avons choisi d'implémenter consiste en un mélange de deux modèles. Chaque modèle comprend alors un filtre principal couvrant approximativement la partie supérieur du corps d'une personne ainsi que six filtres secondaires de plus haute résolution couvrant chacun une composante spécifique du corps. Plus formellement, un modèle de personne contenant n parties est défini par un $(n+2)$-tuple $(F_0, P_1, ..., P_n, b)$ ou F_0 est un filtre principal, P_i est un modèle de la $i^{\text{ème}}$ partie et b correspond au biais. Chaque partie P_i est alors définie par (F_i, v_i, d_i) où F_i est le filtre secondaire correspondant à la $i^{\text{ème}}$ partie, v_i est un vecteur 2D donnant la position relative de la partie i par rapport à la position du filtre principal et d_i correspond aux coefficient d'une fonction quadratique définissant un coût de déformation pour chaque position possible de la partie i par rapport au filtre principal. Dans le cadre de la détection de personne, nous utilisons des histogrammes de gradients orientés [Dalal and Triggs, 2005]. Un filtre F correspond alors à une région d'intérêt définie et sa réponse à une position (u, v) d'une image de gradients orientés G est alors définie par :

$$s(F) = \sum_{x',y'} F[x', y'].G[x + x', y + y'], \tag{2.7}$$

avec (x', y') décrivant l'ensemble des points de F.

Le score final d'une hypothèse faite sur la présence d'une personne est alors donné par l'ensembles des scores de chaque filtre à sa position respective moins un coût de déformation calculé en fonction de la position relative de chaque partie au filtre principal.

Pour de plus amples détails, le lecteur peut se référer à [Felzenszwalb et al., 2009].

(a) (b) (c) (d)

FIG. 2.16 – Exemples de détection de personnes depuis Rackham.

2.4 Conclusion et perspectives

Ce chapitre nous a permis de présenter différentes contributions relatives à la détection et l'identification de personnes en temps réel depuis une plateforme mobile. Ces développements constituent une première brique d'abstraction des capteurs permettant d'extraire de manière automatique et systématique des primitives, visuelles ou RF, utiles pour la détection et l'identification des personnes positionnées aux alentours de la plateforme.

Nous avons, dans un premier temps, évalué différentes méthodes de reconnaissance de visage basées sur des images fixes afin de définir la plus pertinente vis-à-vis du contexte applicatif et des performances souhaitées. Nous avons ainsi décliné des classifieurs usuels et évalué ceux-ci sur des bases d'images directement issues du robot. Notre méthode, basée sur une Analyse en Composante Principale et des Machines à Vecteurs de Support (ou SVM) montre de bonnes performances au regard des autres classifieurs holistiques basés sur l'apparence (*Eigenfaces* et *Fisherfaces*). Les performances relatives aux différents classifieurs étudiés ont été analysées au moyen de courbes ROC qui permettent de définir de manière empirique et coûteuse les meilleurs paramètres pour chaque classifieur. Par la suite, les paramètres libres nécessaires à l'exécution de cet algorithme ont été optimisés à l'aide d'un algorithme génétique afin d'obtenir de meilleurs résultats. Ceci constitue le point central dans ces investigations. L'optimisation des différents paramètres libres par un algorithme génétique NSGA-II a clairement permis d'accroître les performances globales du classifieur. Le classifieur retenu permet d'effectuer une reconnaissance de visage en temps-réel correspondant à notre contexte applicatif.

Dans un deuxième temps, la reconnaissance de visages a été complétée avec un détecteur capable d'identifier une personne, même lorsque celle ci ne fait pas face au robot. Ce capteur RFID permet de localiser un (ou plusieurs) badge autour de la plateforme, tant en azimut qu'en distance, grâce à une carte de multiplexage permettant d'adresser séquentiellement 8 antennes. Notre capteur, issu d'un matériel tiers du marché, a été adapté sur notre plateforme mobile comme un capteur omnidirectionnel d'identification de personnes. L'évaluation des performances du capteur, notamment en environnement relativement encombré, a démontré que l'utilisation des RFID pour détecter et localiser le porteur d'un badge est un choix judicieux car il permet de palier aux défauts de la vision. Ceci constitue la seconde contribution du chapitre. Enfin, des investigations en cours visent à réduire l'encombrement du système RF embarqué.

Les deux fonctionnalités précédemment décrites sont complémentaires. En effet, ces détecteurs, bien qu'intermittents, apportent lorsqu'ils sont présents une information pertinente pour le suivi. La fusion de ces deux capteurs permet de combiner la précision de la vision avec l'identification du RFID.

Plusieurs améliorations possibles sont en cours de développement et d'évaluations.

Les développements concernant le système RFID visent principalement l'amélioration de sa compacité. De nouvelles antennes sont en cours de conception afin de correspondre au modèle présenté en section 2.2.5. Le groupe MINC du LAAS-CNRS réalise actuellement la conception d'antennes RFID (figure 2.17).

Plus généralement, un travail de couplage avec les différents détecteurs pour l'analyse spatio-temporelle est en cours. En effet, l'utilisation de fonctionnalités complémentaires permet de combler les lacunes de certaines méthodes. Par exemple, à l'instar de la complémentarité entre l'identification de visage (intermittent) et l'identification RFID (persistant si non occulté) utilisés pour le suivi de l'utilisateur, l'aspect intermittent du détecteur de personnes peut être complété par des détections laser, plus persistantes, au sein d'un algorithme de suivi multi-cibles.

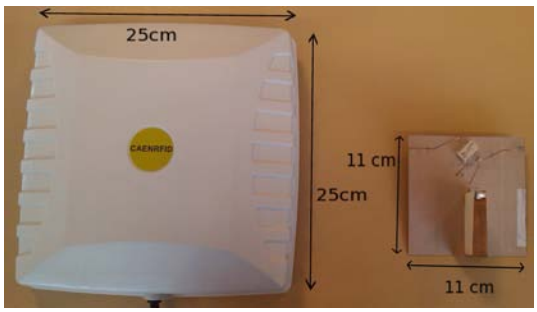


FIG. 2.17 – Antenne du commerce (à gauche) *vs.* nouvelle antenne (à droite).

Chapitre 3

Fusion de données pour le suivi mono-personne

Afin de permettre à un robot mobile d'interagir de manière efficace avec l'homme, il doit, entre autre, pouvoir suivre physiquement cette personne de manière automatique afin de coordonner les mouvements entre le robot et son utilisateur. De nombreuses applications peuvent bénéficier de ces capacités.

Les robots de service ont clairement besoin de bouger de manière coordonnée et sociable pour les personnes. Un tel robot doit pouvoir localiser son utilisateur, l'identifier parmi les passants et être capable de le suivre à travers un environnement encombré et dynamique. Dans ce contexte, un point fondamental est la capacité de suivre une personne donnée dans la foule impliquant une forte réactivité de la part du robot. Cependant, de nombreuses difficultés se posent : mobilité des caméras avec champs de vue limités, arrière plan encombré, variations d'illumination, contraintes temps-réels fortes, etc...

L'interaction Homme / Robot implique donc des mouvements conjoints Homme / Robot. Pour ce faire, il est important pour le robot de conserver, autant que faire se peut, un contact visuel avec sa cible. Les fonctions de détection et identification étudiés au chapitre précédent doivent être complété par une analyse spatio-temporelle offrant au système une vérification continue de la présence ou non de l'utilisateur au voisinage du robot ainsi que de sa distance afin de se mouvoir de façon intelligente par rapport à l'homme. Il existe donc un lien fort entre les notions de détection et de suivi compte tenu du contexte d'application en milieu encombré. En effet, le robot doit être capable de gérer les occultations de la cible, mais aussi de réinitialiser le contact visuel de manière automatique lors de la perte sporadique de la cible dû à une sortie du champ de vue de la cible ou une occultation. Il faut donc être capable d'associer les données issues des différents capteurs de manière robuste au sein d'un système permettant d'estimer la position de l'utilisateur au voisinage du robot. Ce chapitre se concentre sur l'implémentation d'un filtre particulaire permettant la fusion de données hétérogènes et sur les évaluations associées en présence d'artefacts, tels que occultations, disparition de la cible, ..., grâce à des séquences d'images acquises depuis le robot statique. En effet, il est nécessaire de décorreler la fonction sensorielle de suivi de la tâche de guidage physique afin d'évaluer incrémentalement les performances de notre traqueur.

Le chapitre suivant est structuré comme suit. La section 3.1 présente un état de l'art des

différentes stratégies de fusion de données pour la perception de l'homme. La section 3.2 décrit notre approche basée sur le filtrage particulaire dont le formalisme est rappelé en section 3.3. La section 3.4 détaille notre stratégie de fusion de données hétérogènes pour le suivi mono-cible tandis que son implémentation est présentée en section 3.5. Des évaluations qualitatives et quantitatives sont présentées et commentées dans la section 3.6. La section 3.7 résume nos contributions sur ce chapitre ainsi que les perspectives sur notre traqueur multimodal de personne.

3.1 Etat de l'art

Dans la plupart des applications actuelles, la caméra reste le capteur principal par la richesse d'information délivrée. Cependant, la forte dépendance de la vision au contexte de prise de vue (changement d'apparence du modèle, condition d'illumination) induit souvent un couplage avec des données issues de capteurs moins riches, mais aussi plus robustes à ces artefacts.

Dans la problématique énoncée, la littérature propose différentes stratégies de fusion hétérogène de flux perceptuels. En effet, certaines approches visuelles, dédiées généralement aux robots à l'arrêt, consistent à segmenter les personnes en mouvement du fond [Tsai et al., 2006; Zajdel et al., 2005]. Certains travaux [Calisi et al., 2007; Gavrila and Munder, 2007; Nickel et al., 2005] considèrent une segmentation d'arrière-plan basée sur des cartes de disparité issues d'une tête stéréo [Muñoz Salinas et al., 2008], mais ceci nécessite généralement beaucoup de ressources CPU. D'autres techniques supposent que les personnes regardent en direction du robot. Ici, une détection de visage [Bellotto and Hu, 2006; Huang et al., 2007; Viola and Jones, 2003] est appliquée pour (ré-)initialiser de manière efficace un suivi de personne après une occulation temporaire, une sortie du champ de vue ou une perte de cible. Ces détecteurs de visages multi-vues sont largement exploités dans ce contexte. Des approches complémentaires combinent la détection et la reconnaissance de personnes (cf. chapitre 2) [Zajdel et al., 2005] afin de différencier la personne suivie des autres détections. Malgré tout, les fortes variabilités en termes d'illumination, d'orientation de visages et de distance Homme / Robot limitent les performances du suivi. De plus, la détection de visage ainsi que la détection de couleur peau ne sont efficaces que lorsque la personne fait face au robot. Dans ces conditions, il est difficile pour le robot de suivre la cible quelle que soit la situation relative Homme / Robot.

Par la réception de rayonnements infra-rouges, la vision thermique permet de s'affranchir de certaines de ces limitations, puisque l'homme à un profil thermique différents des autres objets inanimés. De plus leur apparence thermique ne dépend pas des conditions d'illumination. Jusqu'à présent, il existe très peu de travaux traitant de la fusion de données thermiques et visibles depuis un robot mobile pour suivre une personne (c.f. le survey de [Hammoud and Davis, 2007]). Nous pouvons ici mentionner les travaux de Cielniak *et al.* [Cielniak et al., 2007] qui utilisent le spectre infra-rouge pour détecter les personnes et le spectre visible pour capter l'apparence. Malheureusement, dans une foule, la perception depuis une caméra thermique est perturbée par un nombre conséquent de personnes présentes dans le champ de vue. En effet, la vision thermique ne permet pas de différencier et d'identifier les cibles présentes, ce qui pose le problème de l'association des données entre les cibles. Il est alors impossible d'identifier une personne

précise du fait que toutes les personnes donnent lieu à la même signature thermique dans l'image infra-rouge.

D'autres systèmes de perception multimodale dédiés au suivi de personne utilisent des capteurs visio-auditifs [Bregonzio et al., 2007; Bohme et al., 2003; Nickel et al., 2005; Pérez et al., 2004]. Dans une foule, le problème d'association de données peut être réglé par l'identification du locuteur [Kar et al., 2007; Ying et al., 2004]. Néanmoins, percevoir des personnes au travers d'attributs auditifs pendant le déplacement du robot ou de la personne est assez complexe. De plus, la variabilité dans l'intonation de la voix, les conditions d'enregistrement, le bruit ambiant de la foule ainsi que l'intermittence de la voix sont autant de limitations à surmonter. L'identification du locuteur reste cependant un sujet ouvert à de nombreuses investigations.

De nombreuses approches de fusion de données utilisent les informations issues d'un laser ainsi que d'une caméra perspective [Bellotto and Hu, 2006; Cui et al., 2008; Spinello et al., 2008] ou omnidirectionnelle [Kobilarov et al., 2006; Zivkovic and Kröse, 2007]. L'avantage d'une telle approche est de combiner des informations géométriques et d'apparence visuelle. Belotto *et al.* [Bellotto and Hu, 2006] utilisent la reconnaissance de visage sur les positions laser des jambes afin d'identifier les personnes. Dans [Cui et al., 2008], la détection des jambes permet d'initialiser un algorithme de suivi visuel et les informations résultant des deux capteurs sont combinées grâce à une méthode de fusion Bayesienne. Zivkovic *et al.* [Zivkovic and Kröse, 2007] introduisent une représentation multimodale basées sur les parties du corps humain. Cette approche combine un détecteur de jambes basé sur le laser (grâce à une méthode de Boost) avec des détecteurs de différentes parties du corps depuis une caméra omnidirectionnelle.
Dans tous les cas, les systèmes qui impliquent des coupes laser souffrent de nombreuses limitations. La détection des jambes dans une coupe 2D ne permet pas de discriminer différentes personnes de manière robuste et la détection échoue lorsqu'une seule jambe est détectée. De plus, le laser ne permet pas d'identifier une personne parmi d'autres du fait qu'aucune information concernant l'apparence n'est donnée.

De récentes approches de suivi de personnes se focalisent sur les techniques de positionnement en milieu intérieur basées sur les infrastructures réseau sans fil, les ultrasons, infrarouges [Schulz et al., 2003], ou les badges radio-fréquence [Anne et al., 2005; Castano and Rodriguez, 2008; Hahnel et al., 2004; Kanda et al., 2007; Takahashi et al., 2008]. Les signaux Radio-Fréquence (RF) sont largement utilisés car (i) ils possèdent une excellente portée en environnement intérieur, (ii) ils produisent peu d'interférences avec les autres composants radiofréquence. Les applications communes impliquant la technologie RFID [Anne et al., 2005; Castano and Rodriguez, 2008; Kanda et al., 2007; Mori et al., 2004; Schulz et al., 2003] disposent de lecteurs fixes distribués dans l'environnement aussi appelés capteurs ubiquistes. Seuls Schutz *et al.* [Schulz et al., 2003] combinent un réseau de capteurs RF et de laser placés dans l'environnement pour le suivi multimodal de personnes.

Il est à noter que la majeure partie des travaux cités ci-dessus impliquent une fusion de données multimodale dans le cadre des détections, *i.e.* en amont d'une quelconque stratégie de filtrage, et non dans la boucle de suivi. A notre connaissance, très peu de travaux traitent d'une analyse spatio-temporelle multimodale hétérogène [Schulz et al., 2003; Pérez et al., 2004]. En

effet, la plupart des travaux sur le suivi de personnes sont basés sur des données homogènes. Il paraît donc intéressant de proposer une stratégie de suivi multimodale fusionnant des données hétérogènes et bénéficiant ainsi des spécificités de chacun des canaux de perception. Dans cette optique, l'utilisation d'une stratégie de filtrage particulaire semble être un choix judicieux pour fusionner diverses sources de données dans un cadre probabiliste justifié.

Rappelons que notre approche privilégie des ressources perceptuelles embarquées (vision couleur monoculaire et signaux RF) afin de limiter l'instrumentation de l'environnement.

3.2 Notre approche

Notre algorithme multimodal de suivi de personne a pour but de combiner la précision et la richesse d'information de la vision couleur avec l'identification du RFID. Notre système étant un système passif, l'impact sur l'appareillage de l'environnement et de l'utilisateur s'en trouve grandement réduit. A notre connaissance, la fusion de données vision / RFID n'a jamais été traité dans la littérature alors que leur combinaison semble prometteuse.

Cette stratégie de suivi multimodal par fusion des signaux vidéo et RF, doit être plus robuste aux occultations qu'une stratégie uniquement basé sur la vision puisqu'il bénéficie d'une orientation approximative de la personne et de son identité en plus de connaître son apparence. De plus, le badge RF peut agir comme un stimuli fort afin de piloter la vision via les actionneurs d'une platine ou du robot. Aussi, lorsque plusieurs personnes sont présentes dans le champ de vue de la caméra[1], la fusion de données multimodale doit permettre de distinguer la personne-cible des autres. Notre objectif est de suivre une personne cible tout au long d'une séquence vidéo. Pour cela, nous cherchons à estimer ses coordonnées (u, v) et son facteur d'échelle s dans le plan image. Tous ces paramètres sont recensés dans un vecteur d'état $\mathbf{x}_k$ à l'instant k.

La fusion de données est largement évoquée dans [Pérez et al., 2004] où Pérez *et al.* mettent en avant le fait que les attributs intermittents contribuent efficacement à la mise en place de détecteurs et de leurs distributions associées. Les données issues de l'identification de visage présentées en section 2.1 correspondent à ces attributs intermittents. Au delà des attributs classiques basés sur l'apparence, nous proposons d'utiliser d'autres attributs, visuels ou non, afin de faciliter le suivi. Notre stratégie propose de combiner au sein de la fonction d'importance la dynamique du système avec certaines mesures comme l'identification visuelle ou RF. De plus, au sein du filtre, la vraisemblance de chaque hypothèse est calculée au moyen de fonctions de mesure utilisant des attribus persistants. L'apparence colorimétrique, attribut persistant, est connue pour grandement améliorer la robustesse du suivi, plus spécialement à faible résolution, lorsque les détails du visage ne peuvent être utilisés. De notre avis, la fusion simultanée d'attributs multiples ne permet pas seulement d'utiliser une information complémentaire et redondante, mais aussi de rendre plus robuste le suivi de personne et la réinitialisation automatique de cibles. Il est donc judicieux de pouvoir échantillonner les particules du filtre suivant une fonction d'importance représentant au mieux la variabilité et la complémentarité des sources. Une telle stratégie basée sur l'exploitation de données hétérogènes n'a été que très peu exploitée dans le cadre du

[1] Dans ce cas, il y a de nombreuses observations sur le plan image.

suivi de personnes [Isard and Blake, 1998b; Pérez et al., 2004]. Ces contributions ont montré que le filtrage particulaire est particulièrement adapté à la fusion de données.

De nombreuses approches de la fusion de données considérent une fusion dans la fonction de vraisemblance [Pérez et al., 2004; Li et al., 2006; Muñoz Salinas et al., 2008; Chateau et al., 2009; Yang et al., 2010] alors que très peu utilisent la fusion de données dans la fonction d'importance pour guider l'échantillonnage des particlues [Jin, 2009]. Le principe a été initié par Isard *et al.* [Isard and Blake, 1998b] dans une stratégie de filtrage particulaire appelée ICONDENSATION. Pour cela, nous présentons ici une fonction d'importance basée sur des cartes de saillance probabilistes, issues des différents détecteurs de personnes, visuels et RF, ainsi qu'un mécanisme d'échantillonnage par rejet afin de (re-)positionner les particules dans les zones pertinentes de l'espace d'état lors du suivi. L'utilisation d'une fonction d'importance combinant au sein d'une même carte de saillance, les modèles de dynamiques de la cible aux techniques d'identification présentées précédemment constitue un apport indéniable quant à l'efficacité de l'échantillonnage des particules. Son couplage avec une stratégie d'échantillonnage des particules par rejet (ou *rejection sampling*), unique dans la littérature, permet d'améliorer les performances de notre algorithme de suivi multimodal en terme de sensibilité aux occultations, aux mauvaises associations de données et aux pertes temporaires de cibles par rapport à un système uniquement basé sur la vision.

3.3 Généralités sur le filtrage particulaire et la fusion de données

Rappelons tout d'abord le principe du filtrage particulaire au travers des stratégies les plus répandues, *i.e.* SIR[2], CONDENSATION[3] et ICONDENSATION[4].

Les techniques de filtrage particulaire sont des méthodes de simulation séquentielles de type Monte Carlo permettant l'estimation du vecteur d'état d'un système Markovien non linéaire soumis à des excitations aléatoires possiblement non Gaussiennes [Arulampalam et al., 2002; Doucet et al., 2001]. En tant qu'estimateurs Bayésiens, leur but est d'estimer récursivement la densité de probabilité *a posteriori* $p(\mathbf{x}_k|z_{1:k})$ du vecteur d'état $\mathbf{x}_k$ à l'instant k conditionné sur l'ensemble des mesures $z_{1:k} = z_1, \ldots, z_k$, une connaissance *a priori* de la distribution du vecteur d'état initial $\mathbf{x}_0$ pouvant être également prise en compte. A chaque instant image k, la densité $p(\mathbf{x}_k|z_{1:k})$ est approximée au moyen de la distribution ponctuelle

$$p(\mathbf{x}_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}), \ \sum_{i=1}^{N} w_k^{(i)} = 1, \tag{3.1}$$

exprimant la sélection d'une valeur – ou particule – $\mathbf{x}_k^{(i)}$ avec la probabilité – ou poids – $w_k^{(i)}$ où $i = 1, \ldots, N$ est l'index de la particule. Les moments conditionnels de $\mathbf{x}_k$, tels que l'estimateur

[2]Pour *Sampling Importance Resampling*.
[3]Pour *Conditional Density Propagation*.
[4]Pour *Independent Conditional Density Propagation*.

du minimum d'erreur quadratique moyenne (ou MMSE, pour *Minimum Mean Square Error*) $\mathrm{E}[\mathbf{x}_k|z_{1:k}] = \sum_{i=1}^{N} w_k^{(i)}\mathbf{x}_k^{(i)}$, peuvent alors être approchés par ceux de la variable aléatoire ponctuelle de densité de probabilité (3.1). Ainsi, nos différents filtres sont basés sur cet estimateur MMSE.

Les particules $\mathbf{x}_k^{(i)}$ évoluent stochastiquement dans le temps. Elles sont échantillonnées selon une fonction d'importance visant à explorer adaptativement les zones "pertinentes" de l'espace d'état.

3.3.1 Algorithme générique ou SIR

L'algorithme SIR (pour *Sampling Importance Resampling*), présenté par l'algorithme 3.1, est entièrement décrit par une connaissance *a priori* $p(\mathbf{x}_0)$, sa dynamique $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ ainsi que ses observations $p(z_k|\mathbf{x}_k)$. Son initialisation consiste en la définition d'un ensemble de particules pondérées décrivant la distribution *a priori* $p(\mathbf{x}_0)$, e.g. en affectant des poids identiques $\{w_0^{(i)} = \frac{1}{N}\}_{i=0}^{N}$ à des échantillons $\mathbf{x}_0^{(1)}, \ldots, x_0^{(N)}$ indépendamment et identiquement distribués (ou i.i.d.) selon $p(\mathbf{x}_0)$.

A chaque instant k, disposant de la mesure z_k et de la description particulaire $\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}$ de $p(\mathbf{x}_{k-1}|z_{1:k-1})$, la détermination de l'ensemble des particules pondérées $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}$ associé à la densité *a posteriori* $p(\mathbf{x}_k|z_{1:k})$ se fait en deux étapes. Dans un premier temps, les $\mathbf{x}_k^{(i)}$ sont échantillonnés selon la fonction d'importance $q(\mathbf{x}_k|\mathbf{x}_{k-1}, z_k)$ évaluée en $\mathbf{x}_{k-1} = \mathbf{x}_{k-1}^{(i)}$, cf. l'équation 3.2. Les poids $w_k^{(i)}$ sont ensuite mis à jour de façon à assurer la cohérence de l'approximation (3.1). Ce calcul obéit à l'équation 3.3, où $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ rend compte de la dynamique du processus d'état sous-jacent, et la vraisemblance $p(z_k|\mathbf{x}_k)$ d'un état possible $\mathbf{x}_k$ vis à vis de la mesure z_k est évaluée à partir de la densité de probabilité relative au lien état-observation.

Toute méthode de simulation séquentielle de type Monte Carlo souffre du phénomène de dégénérescence, au sens où après quelques itérations, les poids non négligeables tendent à se concentrer sur une seule particule. Afin de limiter ce phénomène, une étape de rééchantillonnage, introduite par Gordon *et al.* dans [Gordon et al., 1993], peut être insérée en fin de chaque cycle de l'algorithme SIR (cf. étape 11 de l'algorithme 3.1). Ainsi, N nouvelles particules $\tilde{\mathbf{x}}_k^i$ sont obtenues par rééchantillonnage avec remise dans l'ensemble $\{\mathbf{x}_k^j\}$ selon la loi $P(\tilde{\mathbf{x}}_k^i = \mathbf{x}_k^j) = w_k^j$. Les particules associées à des poids w_k^j élevés sont dupliquées, au détriment de celles faiblement pondérées qui disparaissent, de sorte que la séquence $\tilde{\mathbf{x}}_k^1, \ldots, \tilde{\mathbf{x}}_k^N$ est i.i.d. au regard de (3.1).

Cette étape de redistribution peut soit être appliquée systématiquement, soit être déclenchée seulement lorsqu'un critère d'efficacité du filtre passe en deçà d'un certain seuil [Doucet et al., 2000; Arulampalam et al., 2002]. Le calcul des moments de (3.1) doit de préférence faire intervenir l'ensemble des particules pondérées avant rééchantillonnage.

En complément de la fonction d'importance, la fonction de mesure nécessite l'utilisation d'indices visuels persistants et, par conséquent, plus sujets aux ambiguïtés dans les scènes encombrées. Une alternative peut être de considérer une fusion multi-attributs lors de la mise à jour

ALG. **3.1** Algorithme de filtrage particulaire générique (SIR).

ENTRÉES: $[\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^{N}, z_k$

si $k = 0$ **alors**
  Echantillonner $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(N)}$ i.i.d. selon $p(\mathbf{x}_0)$, et poser $w_0^{(i)} = \frac{1}{N}$, $i = 1, \ldots, N$
fin si
si $k \geq 1$ **alors** —*Soit* $[\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^{N}$ *l'ensemble des particules de* $p(\mathbf{x}_{k-1}|z_{1:k-1})$—
  pour $i = 1, \ldots, N$ **faire**
    « Propager » la particule $\mathbf{x}_{k-1}^{(i)}$ en simulant de manière indépendante

$$\mathbf{x}_k^{(i)} \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, z_k) \tag{3.2}$$

    Mettre à jour le poids $w_k^{(i)}$ associé à $\mathbf{x}_k^{(i)}$ suivant

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k|\mathbf{x}_k^{(i)}) p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{q(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, z_k)} \tag{3.3}$$

  fin pour
  Procéder à une étape de normalisation telle que $\sum_i w_k^{(i)} = 1$
  Calculer le moment conditionnel de $\mathbf{x}_k$, *e.g.* l'estimé MMSE $\mathrm{E}_{p(\mathbf{x}_k|z_{1:k})}[\mathbf{x}_k]$, à partir de l'approximation $\sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)})$ de $p(\mathbf{x}_k|z_{1:k})$
  Rééchantillonner $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}$ selon $P\big(\tilde{\mathbf{x}}_k^{(i)} = \mathbf{x}_k^{(j)}\big) = w_k^{(j)}$, ce qui conduit à un ensemble de particules pondérées $\{\tilde{\mathbf{x}}_k^{(i)}, \frac{1}{N}\}$ tel que $\sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)})$ et $\frac{1}{N} \sum_{i=1}^{N} \delta(\mathbf{x}_k - \tilde{\mathbf{x}}_k^{(i)})$ approximent $p(\mathbf{x}_k|z_{1:k})$
  Affecter $\mathbf{x}_k^{(i)}$ et $w_k^{(i)}$ avec $\tilde{\mathbf{x}}_k^{(i)}$ et $\frac{1}{N}$
fin si

des poids. Soit L_m sources de mesures $(z_k^1, \ldots, z_k^{L_m})$ mutuellement indépendantes, la fonction de mesures unifiée peut être factorisée de la manière suivante :

$$p(z_k^1, \ldots, z_k^{L_m}|\mathbf{x}_k^{(i)}) \propto \prod_{l=1}^{L_m} p(z_k^l|\mathbf{x}_k^{(i)}). \tag{3.4}$$

3.3.2 Echantillonnage guidé par la dynamique ou CONDENSATION

L'algorithme de CONDENSATION [Isard and Blake, 1998a] (pour *Conditional Density Propagation*) est dérivé du SIR par le fait que les particules sont échantillonées suivant la dynamique du système, c'est à dire $q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, z_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})$. Dans le domaine du suivi visuel, l'algorithme original [Isard and Blake, 1998a] défini la vraissemblance de chaque particule à partir de primitives visuelles basées sur les contours. D'autres indices visuels ont aussi été exploités [Pérez et al., 2004]. Dans ces conditions, l'échantillonage des particules peut être sujet à

une perte de diversité dans l'exploration de l'espace d'état. La fonction d'importance $q(.)$ doit alors être défini avec précaution. Or, l'algorithme de CONDENSATION échantillonne les particules $\mathbf{x}_k^{(i)}$ suivant la dynamique du système sans tenir compte des mesures z_k donc, certaines particules peuvent obtenir une faible vraissemblance $p(z_k|\mathbf{x}_k^{(i)})$ et par conséquent un poids faible lors de l'étape 7 de mise à jour de l'algorithme 3.1, ce qui peut entraîner une baisse significative des performances.

3.3.3 Echantillonnage guidé par la mesure ou ICONDENSATION

Le rééchantillonnage utilisé seul ne suffit pas à limiter efficacement le phénomène de dégénérescence évoqué précédemment. En outre, il peut conduire à une perte de diversité dans l'exploration de l'espace d'état, du fait que la description particulaire de la densité *a posteriori* risque de contenir de nombreuses particules identiques. La définition de la fonction d'importance $q(\mathbf{x}_t|\mathbf{x}_{t-1}, z_t)$ – selon laquelle les particules sont distribuées – doit donc également faire l'objet d'une attention particulière [Arulampalam et al., 2002].

En suivi visuel, les modes des fonctions de vraisemblance $p(z_k|\mathbf{x}_k)$ relativement à $\mathbf{x}_k$ sont généralement très marqués. Il s'en suit que les performances sont souvent assez médiocres pour l'algorithme de CONDENSATION. Du fait que les particules sont positionnées selon la dynamique du processus d'état et "en aveugle" par rapport à la mesure z_k, un sous-ensemble important d'entre elles peut être affecté d'une vraisemblance très faible par l'équation $w_k^{(i)} \propto w_{k-1}^i p(z_k|\mathbf{x}_k^{(i)})$, dégradant ainsi significativement les performances de l'estimateur.

Une alternative, appelée MSIR (pour *Measurement-based* SIR), consiste à échantillonner les particules à l'instant k suivant une fonction d'importance $\pi(\mathbf{x}_k|z_k)$ définie sur les mesures courrantes. La première stratégie MSIR était l'algorithme de ICONDENSATION [Isard and Blake, 1998b], qui effectuait l'échantillonnage suivant des détections de blobs de couleur. D'autres fonctionnalités de détections visuelles peuvent aussi être utilisées, *e.g.* la détection/reconnaissance de visages, ou toute autre primitive intermittente qui, malgré un aspect sporadique, peut être très discriminante lorsqu'elle est présente [Pérez et al., 2004]. La fonction d'importance classique $\pi(\mathbf{x}_k|z_k)$ peut alors être étendue afin de considérer L_d différentes mesures, *i.e.*

$$\pi(\mathbf{x}_k^{(i)}|z_k^{1...L_d}) = \sum_{l=1}^{L_d} \kappa_l \pi(\mathbf{x}_k^{(i)}|z_k^l), \text{ avec } \sum \kappa_l = 1 \tag{3.5}$$

où $\pi(\mathbf{x}_k|z_k^l)$ est la fonction d'importance relative au détecteur z_k^l et κ_l est le coefficient de pondération de la mesure z_k^l.

Au sein d'un tel algorithme, si une particule $\mathbf{x}_k^{(i)}$, exclusivement échantillonnée suivant les mesures $\pi(.)$, est inconsistante avec son prédécesseur $\mathbf{x}_{k-1}^{(i)}$ du point de vue de la dynamique, la mise à jour de son poids $w_k^{(i)}$ suivant (3.3) donnera une valeur faible. Une alternative est de définir la fonction d'importance $q(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, z_k)$ comme une somme pondérée de fonctions d'importance $\pi(.)$ et $p(.)$ respectivement basées sur les mesures et sur la dynamique ainsi que

d'une connaissance *a priori* p_0, soit :

$$q(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, z_k) = \alpha\pi(\mathbf{x}_k^{(i)}|z_k) + \beta p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}) + (1-\alpha-\beta)p_0(\mathbf{x}_k) \tag{3.6}$$

avec $\alpha, \beta \in [0; 1]$.

Ainsi, les différentes mesures contribuent pour α à l'élaboration de la fonction d'importance, la dynamique y contribue pour β. Le reste correspond à une connaissance *a priori*. De ce fait, une détection permet de (ré)initialiser le filtre. De même, en cas de fausses mesures/détections, ou en cas d'absence de mesures, les particules continuent à évoluer selon la dynamique et la connaissance *a priori* du système permettant alors de conserver une représentation optimale de la distribution *a posteriori* $p(\mathbf{x}_k|z_{1:k})$.

3.4 Fonction d'importance et fusion de données

Un problème connu pour tout type de traqueur est la perte de la cible au cours du suivi. Au sein du filtre à particules, la fonction d'importance $q(.)$ (équation 3.6) a pour but de diriger l'échantillonnage des particules dans les zones pertinentes de l'espace d'état de $\mathbf{x}_k$. Il faut donc définir un mécanisme permettant de réinitialiser le filtre de manière automatique sur la cible. En d'autres termes, il faut échantillonner les particules dans la zone de l'espace d'état correspondant à la cible, au moyen de différents détecteurs $\pi(\mathbf{x}_k|z_k^l)$.

3.4.1 Description et prototypage de la fonction $q(.)$

Rappelons que la fonction $\pi(\mathbf{x}_k|z_k)$ dans l'équation 3.5 permet de guider l'échantillonnage des particules grâce à la fusion de données multiples , possiblement hétérogènes. Le choix de l'ensemble des détecteurs $\{\pi(\mathbf{x}_k|z_k^l)\}_{l=1}^{L_d}$ doit donc permettre de positionner efficacement les particules dans l'espace d'état de $\mathbf{x}_k$.

Le chapitre 2 nous a permis de développer et de valider la pertinence de différentes fonctions de détection et d'identification de personnes dans un contexte robotique. Rappelons alors que nous avons défini : (i) une méthode d'identification visuelle de personne permettant d'obtenir, pour tout visage détecté $\mathcal{F}$, une vraissemblance $P(C|\mathcal{F}, z)$ (section 2.1), (ii) un modèle d'observation gaussien, $(\mu_\theta^{tag}, \sigma_\theta^{tag})$ et $(\mu_\rho^{tag}, \sigma_\rho^{tag})$, basé sur des détections RFID (section 2.2).

Nous considérons ici trois fonctions $\pi(\mathbf{x}_k|z_k^c)$, $\pi(\mathbf{x}_k|z_k^s)$ et $\pi(\mathbf{x}_k|z_k^r)$, respectivement basées sur une image de probabilité peau, une identification de visages et une identification RFID.

La première fonction d'importance $\pi(\mathbf{x}_k|z_k^c)$ est basée sur la détection dans l'image z_k^c des zones de couleur chair (peau). Pour celà, la rétro-projection d'un histogramme représentant une distribution de couleur peau dans l'image z_k^c est utilisée [Lee et al., 2003]. La fonction d'importance $\pi(\mathbf{x}_k|z_k^c)$ en $\mathbf{x}_k = (u, v)$ est alors décrite par

$$\pi(\mathbf{x}|z^c) = \mathbf{h}(c_z(\mathbf{x})) \tag{3.7}$$

où $c_z(\mathbf{x})$ est la couleur du pixel situé en $\mathbf{x}$ dans l'image d'origine z^c et $\mathbf{h}$ est l'histogramme 3D normalisé, indexé par les canaux R, G, B représentant la distribution de couleur peau apprise *a priori*.

La seconde fonction $\pi(\mathbf{x}_k|z_k^s)$ est basée sur une image de probabilité construite à partir du détecteur de visage décrit par Viola *et al.* dans [Viola and Jones, 2003]. Soit N_s le nombre de visages détectés et $\mathbf{p}_i = (u_i, v_i)$ le centre de la région i correspondant au visage $\mathcal{F}_i$. La fonction $\pi(\mathbf{x}_k|z_k^s)$ en $\mathbf{x}_k = (u, v)$ est décrite en tant que mélange de gaussiennes comme suit :

$$\pi(\mathbf{x}|z^s) \propto \sum_{j=1}^{N_s} P(C|\mathcal{F}_j, z).\mathcal{N}(\mathbf{x}; \mathbf{p}_j, \mathrm{diag}(\sigma_{u_j}^2, \sigma_{v_j}^2)), \tag{3.8}$$

avec $P(C|\mathcal{F}_j, z)$ la probabilité de vraisemblance entre le visage détecté $\mathcal{F}_j$ et celui de la personne suivie, appris *a priori* suivant la méthode décrite section 2.1. Les variables σ_{u_j} et σ_{v_j} dépendent respectivement de la largeur et de la hauteur du visage $\mathcal{F}_j$ détecté.

Enfin, la troisième fonction $\pi(\mathbf{x}_k|z_k^r)$ repose sur les données RF. Elle est définie par projection de la position du badge RFID, représentée par $(\mu_\theta^{tag}, \sigma_\theta^{tag})$, dans le plan image. La fonction d'importance $\pi(\mathbf{x}_k|z_k^r)$ en $\mathbf{x}_k = (u, v)$ suit la loi suivante :

$$\pi(\mathbf{x}|z^r) = \mathcal{N}(\theta_\mathbf{x}; \mu_\theta^{tag}, \sigma_\theta^{tag}), \tag{3.9}$$

où $\theta_\mathbf{x}$ est la position en azimut de $\mathbf{x}$ dans le repère du robot, déduite de sa position horizontale dans l'image u et de l'orientation de la caméra dans le repère du robot. Les variables μ_θ^{tag} et σ_θ^{tag}, décrites dans la section 2.2, sont respectivement la moyenne et l'écart-type de la position estimée du badge RFID associé à l'utilisateur dans le repère du robot.

La fonction d'importance généralisée s'écrit donc, d'après (3.5) :

$$\pi(\mathbf{x}_k|z_k^c, z_k^s, z_k^r) = \kappa_c\pi(\mathbf{x}_k|z_k^c) + \kappa_s\pi(\mathbf{x}_k|z_k^s) + \kappa_r\pi(\mathbf{x}_k|z_k^r)$$

où κ_l est le coefficient de pondération de la carte de saillance associée à z_k^l. Une illustration de cette fonction d'importance est présentée en figure 3.1(e)

Rappelons que nous cherchons à estimer les paramètres (u, v, s) qui composent le vecteur d'état $\mathbf{x}_k$ à l'instant k. Concernant la dynamique $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, les déplacements d'un humain dans une image sont difficiles à caractériser. Cette faible connaissance est représentée par la définition du vecteur d'état $\mathbf{x}_k = [u_k, v_k, s_k]'$ et l'évolution de ses paramètres suit un modèle indépendant de marche aléatoire $p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{x}_{k-1}, \Sigma)$ où la covariance $\Sigma = \mathrm{diag}(\sigma_u^2, \sigma_v^2, \sigma_s^2)$. La dynamique intervenant dans la fonction d'importance $q(.)$, le paramètre s du vecteur d'état $\mathbf{x}_k$ ne sera échantillonné qu'en fonction de $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, car aucun des détecteurs présentés dans cette section ne permet de guider l'échantillonnage de s de manière précise.

3.4.2 Echantillonnage par rejet

La fonction d'importance $q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, z_k)$ (équation 3.6) est une fonction multimodale résultant de la fusion pondérée (i) de la dynamique du système $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, (ii) de la fonction d'importance $\pi(\mathbf{x}_k|z_k^c, z_k^s, z_k^r)$ associant différents détecteurs, (iii) d'une connaissance *a priori* du système $p_0(\mathbf{x}_k)$. La définition de la fonction $q(.)$ amène donc à considérer un algorithme d'échantillonnage des particule capable de gérer cet aspect multimodal. L'échantillonnage des particules est donc réalisé à l'aide de la fonction d'importance $q(.)$ et d'un algorithme d'échantillonnage par rejet (ou *rejection sampling*).

Le principe est décrit dans l'algorithme 3.2 où $g(.)$ désigne une distribution auxiliaire facilitant l'échantillonnage sous la contrainte que $q(.) \leq Mg(.)$ avec $M \geq 1$ étant une borne supérieure de $\frac{q(.)}{g(.)}$.

ALG. **3.2** Algorithme d'échantillonnage par rejet.

1: **répéter**
2: Tirer $\mathbf{x}_k^{(i)}$ suivant $Mg(\mathbf{x}_k)$
3: $r \leftarrow \dfrac{q(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, z_k)}{Mg(\mathbf{x}_k^{(i)})}$
4: Tirer u suivant $\mathcal{U}_{[0,1]}$
5: **jusqu'à** $u \leq r$

Dans notre cas, $g(.)$ suit une loi uniforme bornée par les limites image telle que $g(.) = \mathcal{U}_{[(0,0),(l,h)]}$ où l et h correspondent à la largeur et la hauteur de l'image et $M = 1$.

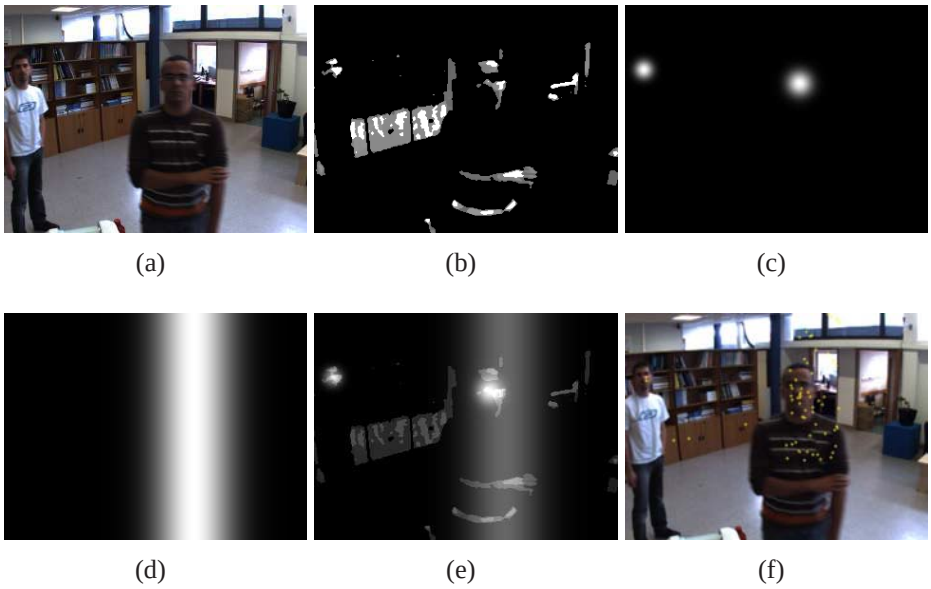
La figure 3.1 montre une illustration de l'algorithme d'échantillonnage par rejet sur une image donnée. Chaque détecteur détaillé précédemment est représenté par une carte de saillance où chaque pixel (u, v) correspond à la réponse du-dit détecteur appliqué sur l'image d'origine (figure 3.1(a)). Les différentes mesures permettent respectivement d'extraire les zones de couleur peau (figure 3.1(b)), les visages (figure 3.1(c)) ainsi que la présence d'un badge RF. Par exemple, la figure 3.1(d) représente la projection de $\pi(\mathbf{x}|z^r)$ dans le plan image *i.e.* la position azimutale du badge RFID détecté. Chaque carte de saillance $\mathcal{I}_k^l$ relative à la fonction d'importance $\pi(\mathbf{x}|z^l)$ est construite telle que :

$$\forall \mathbf{x}_k = (u_k, v_k) \in \mathcal{I}_{z_k},\ \mathcal{I}_k^l = \pi(\mathbf{x}_k|z_k^l),$$

où $\mathcal{I}_{z_k}$ est l'image d'origine sur laquelle est appliquée le détecteur l. Chaque carte de saillance permet de représenter une fonction d'importance qui n'est pas nécessairement définie de manière analytique.

L'ensemble de ces données hétérogènes est alors fusionné au sein de la fonction $q(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, z_k) = \pi(\mathbf{x}_k^{(i)}|z_k^c, z_k^s, z_k^r)$ représentée par la carte de saillance 3.1(e). La figure 3.1(f) montre alors le résultat de l'échantillonnage des particules suivant $q(.)$ d'après l'algorithme d'échantillonnage par rejet.

On peut constater que la majorité des particules se trouve concentrée sur la personne cible située au premier plan, et plus précisément sur son visage, zone combinant les détecteurs peau,



(a) (b) (c)

(d) (e) (f)

FIG. 3.1 – (a) image d'origine, (b) image de probabilité peau (3.7), (c) détection et identification de visages (3.8), (d) détection RFID (3.9), (e) fonction d'importance unifiée (3.6) (sans la dynamique), (f) particules échantillonnées.

visages et RFID. En effet, une zone de l'image de probabilité (figure 3.1(e)) combinant plusieurs attributs possède un mode plus important qu'une zone ne contenant qu'un seul des trois détecteurs décrits ci-dessus. Néanmoins, les zones de moindre importance sont aussi explorées. Ainsi, on peut observer quelques particules situées sur le visage de la deuxième personne, ainsi que sur le mobilier, car celui-ci a une teinte proche de la couleur chair. Ce comportement permet, en cas de perte de la cible, d'explorer l'ensemble des zones de l'image susceptibles de voir réapparaître la personne d'intérêt.

Notre fonction d'importance (3.6) combinée à cette technique d'échantillonnage par rejet assure donc un échantillonnage pertinent des particules sur les zones pertinentes de l'espace d'état. En effet, grâce à cette technique, l'ensemble des particules est échantillonné suivant $q(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, z_k)$, là où la majorité des méthodes échantillonne une proportion α des particules suivant les détections $\pi(\mathbf{x}_k^{(i)}|z_k)$ et une autre proportion β suivant la dynamique $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. Ainsi, chaque hypothèse est affectée identiquement par l'ensemble des paramètres nécessaires à son évolution au cours du temps.

3.5 Implémentation de notre traqueur

Autant dans la fonction d'importance que dans la fonction de mesure, la fusion de multiples attributs permet à l'algorithme de suivi de bénéficier d'informations distinctes et ainsi de réduire sa sensibilité aux mauvaises associations de données. La fonction de vraisemblance impliquée dans la pondération des particules est plus ou moins classique. Elle est basée sur la fusion de

plusieurs fonctions de mesures, liées à des attributs visuels persistants comme (i) une double distribution de couleur représentant l'apparence de l'utilisateur (tête et buste), (ii) un contour pour modéliser la silhouette de la tête.

La fonction de mesure globale $p(z_k|\mathbf{x}_k)$ de notre filtre considère plusieurs régions d'intérêt distinctes spatialement et colorimétriquement [Nummiaro et al., 2003; Pérez et al., 2002], typiquement le visage et les vêtements d'une personne. L'ajout d'une seconde distribution de couleur liée aux vêtements permet la différenciation du sujet guidé lorsque plusieurs individus sont dans le champ de vue. De plus, la gestion de plusieurs sous-régions limite les dérives temporelles observées dans le temps [Pérez et al., 2002]. Ce modèle est relatif aux deux régions d'intérêt rattachées au modèle de la cible. Ces zones sont caractérisées par des distributions locales de couleurs dans l'image. En posant $\mathbf{h}_{\mathbf{x}} = \bigcup_{p=1}^{2} \mathbf{h}_{\mathbf{x},p}$, le modèle de mesure colorimétrique $p(z_k^c|\mathbf{x}_k)$ s'écrit classiquement :

$$p(z_k^c|\mathbf{x}_k) \propto \exp\left(-\sum_c \sum_{p=1}^{2} \frac{D^2(\mathbf{h}_{\mathbf{x},p}^c, \mathbf{h}_{ref,p}^c)}{2\sigma_c^2}\right), \tag{3.10}$$

où $c \in \{R, G, B\}$, σ_c est un écart-type prédéfini, et D est la distance de Bhattacharyya [Aherne et al., 1997] utilisée pour comparer les histogrammes normalisés $\{\mathbf{h}_{\mathbf{x},p}^c\}_{p=1}^2$ et $\{\mathbf{h}_{ref,p}^c\}_{p=1}^2$ respectivement relatifs à l'état $\mathbf{x}$ et au modèle, *i.e.* pour le canal c,

$$D(h_{\mathbf{x}}^c, \mathbf{h}_{ref}^c) = \left(1 - \sum_{j=1}^{N_{bi}} \sqrt{\mathbf{h}_{\mathbf{x}}^c(j).\mathbf{h}_{ref}^c(j)}\right)^{1/2},$$

où N_{bi} correspond au nombre d'intervalles composant l'histogramme $\mathbf{h}_c$.

Enfin, dans notre contexte, les changements d'apparence du sujet observé, de par ses mouvements réels ou les variations d'illumination, impliquent une réactualisation du modèle à chaque instant k. Cette mise à jour est donnée par [Nummiaro et al., 2003] :

$$\mathbf{h}_{ref,k}^c = (1-\kappa).\mathbf{h}_{ref,k-1}^c + \kappa.\mathbf{h}_{E[\mathbf{x}_k]}^c, \tag{3.11}$$

où l'indice p est omis pour plus de clarté et κ pondère l'influence de l'histogramme $\mathbf{h}_{E[\mathbf{x}_k]}$ correspondant à l'état moyen $E[\mathbf{x}_k]$ dans la réactualisation.

Cette mise à jour des distributions de référence $\mathbf{h}_{ref,p}^c$ peut induire des dérives lors du suivi. Ces dérives sont contrôlées par la fusion dans la fonction de mesure globale d'un attribut de forme. Le principe est alors de recaler le modèle sur les contours de la tête avant d'effectuer la mise à jour des distributions. Ainsi, la silhouette de la tête est classiquement modélisée par une spline tandis que les particules sont pondérées à partir des observations constituées des contours image suivant des directions orthogonales à la spline aux points de contrôle [Isard and Blake, 1996]. A l'instant k, le modèle de mesure $p(z_k^s|\mathbf{x}_k)$ s'écrit :

$$p(z_k^s|\mathbf{x}_k) \propto exp\left(-\frac{D^2}{2\sigma_s^2}\right), D = \sum_{l=0}^{N_p} |x(l) - z(l)|, \tag{3.12}$$

où $x(l)$ et $z(l)$ désignent respectivement le $l^{ème}$ point de contrôle sur la spline et le point de contour le plus proche associé sur la normale à la spline.

Supposant indépendantes les mesures z_k^s et z_k^c, d'après l'équation 3.4 la fonction de mesure globale s'écrit :

$$p(z_k^s, z_k^c|\mathbf{x}_k) = p(z_k^s|\mathbf{x}_k).p(z_k^c|\mathbf{x}_k). \tag{3.13}$$

La figure 3.2 illustre la représentation de cette fonction de mesure.

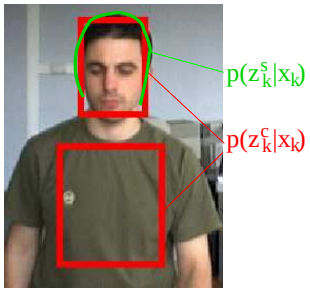


FIG. 3.2 – Modèles de mesure : couleur (en rouge) et forme (en vert).

Notre fonction de mesure est très peu coûteuse au regard de son pouvoir discriminant en terme d'apparence de la personne. L'expression de $p(z_k^s, z_k^c|\mathbf{x}_k)$ reste néanmoins classique et ne constitue pas l'originalité de notre approche. Le réglages de paramètres libres définis dans les sections précédentes (N, $\sigma_{(.)}$) a été étudié dans [Brèthes, 2005] alors que les coefficients de pondération ((α, β), $\kappa_{(.)}$) sont définis empiriquement. Les valeurs de ces paramètres libres correspondant à l'implémentation de notre filtre sont listées table 3.1.

Symbole	Description	Valeur
N	nombre de particules du filtre (3.1)	100
(α, β)	coeff. de la fonction d'importance $q(\mathbf{x}_k|\mathbf{x}_{k-1}, z_k)$ (3.6)	$(0.4, 0.6)$
$(\sigma_u, \sigma_v, \sigma_s)$	écart-type du modèle de marche aléatoire	$(40, 20, 0.2)$
$(\kappa_c, \kappa_s, \kappa_r)$	coeff. de pondération des détecteurs dans $\pi(x_k^{(i)}|z_k^1, \ldots, z_k^L)$ (3.5)	$(0.2, 0.6, 0.2)$
κ	coeff. de mise à jour de l'histogramme de référence $h_{ref,1}^c, h_{ref,2}^c$ (3.11)	0.1
σ_c	dispersion de la vraisemblance couleur $p(z^c|\mathbf{x}_k)$ (3.10)	0.2
σ_s	dispersion de la vraisemblance contour $p(z^s|\mathbf{x}_k)$ (3.12)	20

TAB. 3.1 – Valeurs des paramètres utilisées pour le suivi de personne.

Un dernier point concerne la détection de la présence de la personne cible dans le champ de vision de la caméra. En effet, bien que notre approche permette de gérer les décrochages du filtre avec une réinitialisation automatique *via* les détecteurs, il n'est pas possible de savoir avec certitude si l'estimé MMSE correspond exactement à la personne cible notamment lorsque la cible sort du champ de vue.

Pour celà, nous avons mis en place une heuristique simple permettant de valider la cohérence de l'estimé du filtre. Cette heuristique $\mathcal{H}_{E[\mathbf{x}_k]}$ est basée sur la correspondance entre la position de l'estimé $E[\mathbf{x}_k]$ et celle du badge RFID qui, lorsqu'il est présent, donne une information précise sur l'identité de la cible. Nous posons :

$$\mathcal{H}_{E[\mathbf{x}_k]} = \begin{cases} 0 \text{ si } \pi(E[\mathbf{x}_k]|z_k^r) < \phi \\ 1 \text{ sinon} \end{cases} \tag{3.14}$$

où ϕ est un seuil défini *a priori*.

Le calcul de l'estimé MMSE à l'étape 10 de l'algorithme 3.1 est donc conditionné par $\mathcal{H}_{E[\mathbf{x}_k]}$. En effet, lorsque l'estimé $E[\mathbf{x}_k]$ diffère trop de la position réelle du bagde dans le repère image, il est fort probable que : (1) l'estimation de la position de la cible soit fausse, (2) que la cible soit sortie du champ de vue.

3.6 Evaluations et commentaires associés

(a) Séquence #1 impliquant deux personnes d'apparence vestimentaire différente, de bonnes conditions d'illumination (différence colorimétrique entre le fond et la cible) et de brèves occultations.

(b) Séquence #2 impliquant deux personnes d'apparence vestimentaire différente, des conditions d'illumination difficiles (image saturée par endroits), de brèves occultations et des sorties de la cible du champ de vue. De plus, la couleur du fond et l'apparence de la cible sont similaires.

(c) Séquence #3 impliquant deux personnes d'apparence vestimentaire similaire, de bonnes conditions d'illumination, de brèves occultations et des sorties de la cible du champ de vue. De plus, la couleur du fond et l'apparence de la cible sont similaires.

(d) Séquence #4 impliquant trois personnes dont deux ont la même apparence vestimentaire, de bonnes conditions d'illuminations, des occultations longues et des sorties de la cible du champ de vue. De plus, la couleur du fond et l'apparence de la cible sont similaires.

FIG. 3.3 – Détails des séquences d'images utilisées pour l'évaluation des performances de notre approche. Chacune des séquences comprend (1) plusieurs personnes, (2) occultations de la cible, (3) disparitions du champ de vue de la cible, (4) déplacements erratiques des acteurs.

L'algorithme de suivi a été prototypé sur un processeur Pentium Dual Core 1.8GHz sous Linux à l'aide de la bibliothèque OpenCV. Les évaluations quantitatives et qualitatives hors-ligne sont ici présentées. La base contient diverses séquences, soit un total de plus de 1500 images acquises depuis nos plateformes dans des conditions diverses en terme de situation Homme / Robot (nombre de personnes, distance Homme / Robot, conditions d'illumination, ...). Dans notre contexte, il ne nous est pas possible d'utiliser les bases d'images publiques, telles que [PETS, 2004; PETS, 2006; CLEAR, 2007], qui ne rassemblent pas les conditions de prises de vue et les contraintes inhérentes à notre application. En effet, dans ces bases, les images sont acquises dans des conditions différentes des nôtres, *i.e.* pour des applications de vidéo-surveillance ou pour l'interprétation d'activités humaines. De plus, elles ne comportent pas de données issues d'autres capteurs, *i.e.* données RFID indispensables dans notre stratégie de fusion de données hétérogènes.

Rappelons que nous souhaitons ici évaluer notre stratégie de fusion de données hétérogènes au sein de la fonction d'importance au regard des stratégies de fusion proposées dans la littérature. Ces évaluations seront complétées par des évaluations robotiques en ligne dans le chapitre 5.

Notre base d'images nous permet (i) de déterminer empiriquement les valeurs des paramètres ((α, β), $\kappa_{(.)}$) de l'algorithme, (ii) d'identifier ses forces et ses faiblesses, et en particulier de caractériser sa robustesse aux artefacts de l'environnement *i.e.* encombrement, occultations, sorties du champ de vision, changement d'illumination. Pour chaque séquence, chaque algorithme a été exécuté plusieurs fois afin de s'affranchir de l'aspect stochastique du filtrage particulaire et donc de vérifier la répétabilité du filtre. La figure 3.3 illustre des images clés de chacune des quatre séquences utilisées pour les évaluations. Ces séquences mettent en avant des scènes de plusieurs personnes se déplaçant librement devant le champ de vue de la caméra. Chacune comprend des occultations de la cible, des changements de position et trajectoire des différentes personnes, des disparitions de la cible et des changements d'illumination, relatifs au contexte applicatif.

Les séquences d'images présentées en table 3.2 montrent le comportement qualitatif sur une séquence type (séquence #3, figure 3.3(c)) de notre stratégie de fusion de données, tant au sein de la fonction d'importance (*i.e.* identification de visage, couleur peau, RFID) que dans la fonction de mesure (*i.e.* distribution de couleur, contour). Ces résultats sont commentés ci-après. L'estimé MMSE du filtre de la position de la cible est représenté par les rectangles bleus (double distribution de couleur) et la courbe verte (contour) alors que les points rouges représentent les hypothèses et leurs poids respectifs après normalisation (le noir correspond à un poids $w_k^{(i)} = 0$ et rouge correspond à un poids $w_k^{(i)} = 1$).

La première ligne (table 3.2(a)) montre les résultats de la stratégie $\mathrm{CONDENSATION}$ basée sur (i) l'échantillonnage des particules suivant une dynamique de marche aléatoire et (ii) la mesure de couleur multi-zones. Après quelques itérations, nous observons une dérive du modèle car l'histogramme de référence est alors corrompu par une mise à jour basée sur un arrière-plan encombré. La deuxième ligne (table 3.2(b)) suit mieux la personne cible du fait que la fonction de mesure considère le modèle de contours en plus du modèle de couleur. Même si le modèle suit une personne au lieu de se "perdre" sur l'arrière-plan, la fusion d'attributs au sein de la fonction de mesure n'est pas suffisante pour rester robuste aux occultations entre personnes. En effet, on peut observer un changement de cible entre les images $k = 15$ et $k = 81$. La ligne

Stratégies de fusion de données	$k = 15$.	$k = 81$	$k = 126$	$k = 284$
(a) $q(x_k\|x_{k-1}, z_k) = p(\mathbf{x}_k\|\mathbf{x}_{k-1})$ $p(z_k\|\mathbf{x}_k) = p(z_k^c\|\mathbf{x}_k)$				
(b) $q(x_k\|x_{k-1}, z_k) = p(\mathbf{x}_k\|\mathbf{x}_{k-1})$ $p(z_k\|\mathbf{x}_k) = p(z_k^s\|\mathbf{x}_k).p(z_k^c\|\mathbf{x}_k)$				
(c) $q(x_k\|x_{k-1}, z_k) = \alpha\pi(\mathbf{x}_k\|z_k^c, z_k^s) + \beta p(\mathbf{x}_k\|\mathbf{x}_{k-1})$ avec détection de visages $p(z_k\|\mathbf{x}_k) = p(z_k^s\|\mathbf{x}_k).p(z_k^c\|\mathbf{x}_k)$				
(d) $q(x_k\|x_{k-1}, z_k) = \alpha\pi(\mathbf{x}_k\|z_k^c, z_k^s) + \beta p(\mathbf{x}_k\|\mathbf{x}_{k-1})$ avec identification de visages $p(z_k\|\mathbf{x}_k) = p(z_k^s\|\mathbf{x}_k).p(z_k^c\|\mathbf{x}_k)$				
(e) $q(x_k\|x_{k-1}, z_k) = \alpha\pi(\mathbf{x}_k\|z_k^c, z_k^s, z_k^r) + \beta p(\mathbf{x}_k\|\mathbf{x}_{k-1})$ avec identification de visages et RFID $p(z_k\|\mathbf{x}_k) = p(z_k^s\|\mathbf{x}_k).p(z_k^c\|\mathbf{x}_k)$				

TAB. 3.2 – Cinq stratégies de fusion de données pour l'échantillonnage préférentiel et la fonction de mesure.

(c) de la table 3.2 combine la détection de visages et de pixels de couleur peau avec la dynamique de marche aléatoire au sein de la fonction d'importance afin de diriger l'échantillonnage des particules dans des zones spécifiques de l'image courante (principalement sur des détections de visages). Nous pouvons voir que cette stratégie n'est pas suffisante pour faire une distinction entre l'une ou l'autre des cibles. La ligne (d) de la table 3.2 montre le comportement de l'algorithme de ICONDENSATION lorsque la détection de pixels de couleur peau est fusionnée, non pas avec un détecteur de visage, mais avec l'identification de visages présentée au chapitre 2 (équation 3.8) ainsi qu'avec la dynamique de marche aléatoire dans la fonction d'importance. Nous pouvons voir, à l'image $k = 81$, qu'après une occultation sporadique de la cible par une autre personne (portant un pantalon noir), le processus d'identification de visage aide à repositionner l'échantillonnage des particules uniquement sur la personne cible et aide, par la même occasion à rétablir le contact visuel avec la personne cible. Néanmoins, si la personne cible ne fait pas face à la caméra suite à une occultation, le processus d'identification ne peut aider à recouvrer la cible comme pour les images $k = 126$ et $k = 284$. Dans la ligne (e) de la table 3.2, la fonction d'importance correspond à celle décrite par les équations 3.6 et 3.5 avec $L_d = 3$ et $l_i \in \{c, s, r\}$. On peut alors observer qu'après une occultation, la personne cible ne doit pas nécessairement faire face à la caméra pour que le contact visuel soit rétabli. Ceci est principalement dû à l'utilisateur du détecteur RFID.

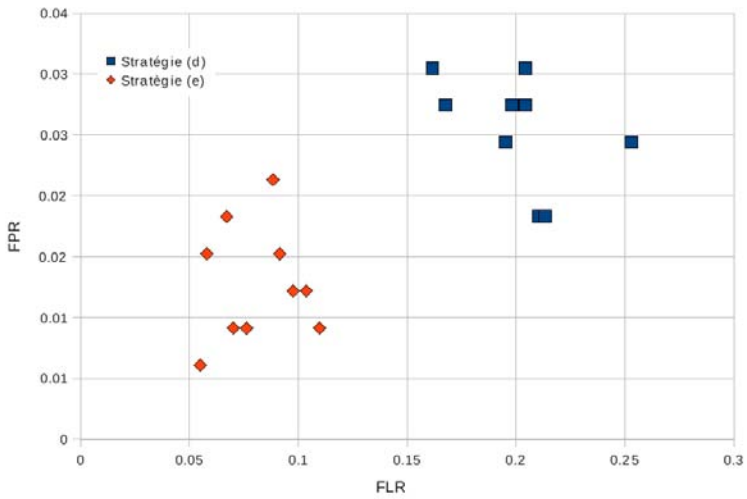


FIG. 3.4 – Performance du suivi de personne pour les traqueurs présentés table 3.2(d-e).

L'évaluation des performances quantitatives de l'algorithme a été réalisée sur notre base d'images illustrée par la figure 3.3. Vu que le point important du suivi est sa robustesse aux artefacts de l'environnement, nous avons comparé les performances du suivi en terme de Taux de Fausses Positions (ou FPR pour *False Position Rate*) et de Taux de Fausses Identification (ou FLR pour *False Label Rate*) respectivement relatifs à la position du modèle dans l'image et l'identification de la bonne cible sont définis comme suit :

$$FPR = \frac{\text{nombre de fausses positions}}{\text{nombre total d'images}} \text{ et } FLR = \frac{\text{nombre de fausses identifications}}{\text{nombre total d'images}}$$

Si le filtre décroche, ceci est considéré comme étant une fausse position alors que si le modèle commute sur une personne autre que la personne cible, une fausse identification sera considérée. La figure 3.4 montre les performances de notre système de fusion multimodale de données hétérogènes (table 3.2(e)) par rapport à une stratégie de fusion de données plus classique (table 3.2(d)) sur la séquence $\#2$.

Les différents points représentent les résultats observés sur 10 réalisations. Il apparaît clairement que notre système visant à combiner les données issues de différents capteurs complémentaires permet de réduire considérablement les erreurs d'identification dues à une occultation ou à une disparition de la cible. En effet, le nombre moyen de fausses identifications est réduit de 12% ($0.08\ vs.\ 0.20$) et, par conséquent, l'estimation de la position de la cible est plus précise lorsque la fusion hétérogène est utilisée. En effet, le FPR moyen est, lui aussi, réduit de 12%($0.13\ vs.\ 0.25$).

Enfin, pour évaluer la precision du filtre, nous avons représenté sur la figure 3.5, l'évolution de la distance moyenne entre l'estimé MMSE et la vérité-terrain sur 10 réalisations de la séquence présentée en figure 3.6 (séquence #2, figure 3.3(b)). La vérité-terrain (ou GT pour *Ground Truth*), établie manuellement sur l'ensemble de la séquence, est définie à l'instant k par $\mathbf{x}_k^{GT} = (u_k^{GT}, v_k^{GT}, s_k^{GT})'$. Il est alors possible de définir l'erreur à l'instant k telle que :

$$\mathcal{E}_k = \mathcal{D}(E[\mathbf{x}_k], \mathbf{x}_k^{GT}) = \sqrt{(E[\mathbf{x}_k] - \mathbf{x}_k^{GT})'.(E[\mathbf{x}_k] - \mathbf{x}_k^{GT})}$$

La courbe rouge représente la présence effective de la personne cible dans l'image. Nous pouvons alors observer certains pics sur cette courbe. La majeure partie d'entre eux est due (1) à des occultations temporaires de la cible ($t = 61, 62, 63; t = 101, 102, 103; t = 128, 129, 130$), (2) à des disparitions du champ de vue ($t = 21, \ldots, 26; t = 77, \ldots, 83; t = 156, \ldots$). De plus, on peut aussi observer un brève dérive du filtre lorsque la cible se trouve dos à la caméra ($t = 68, \ldots, 71, t = 92, \ldots, 97$). Le filtre n'utilise alors que l'information fournie par le badge RFID qui reste moins précise au regard de la position que la vision. Néanmoins, ces dérives sont rapidement corrigées par une détection et les performances globales de notre système restent suffisamment bonnes au regard de l'application choisie.

La figure 3.7 montre des images-clés extraites de la séquence de test $\#4$ ainsi que les résultats de notre stratégie de fusion associés. Cette séquence met en situation un utilisateur et deux passants, comporte de nombreuses occultations ainsi que deux pertes de contact avec l'utilisateur. L'ensemble des concepts présentés dans ce chapitre sont ici mis en œuvre tels que (i) la fusion de données dans la fonction d'importance $q(\mathbf{x}_k|\mathbf{x}_{k-1}, z_k)$ (équation 3.6), (ii) la fusion de donnée dans la fonction de vraisemblance $p(z_k^s, z_k^c|\mathbf{x}_k)$ (équation 3.13), (iii) l'heuristique $\mathcal{H}_{E[\mathbf{x}_k]}$ (équation 3.14). Les cartes de saillance globales (colonne 3.7(b)) associées à chaque itération (colonne 3.7(a)) permettent alors l'échantillonnage des particules en fonction des différentes détecteurs présents. La colonne 3.7(c) montre alors l'état du nuage des particules aprés échantillonnage et pondération des particules par la fonction de vraisemblance, alors que la colonne 3.7(d) donne le résultat de l'estimé MMSE ainsi que l'état du nuage des particules aprés l'étape de rééchantillonnage.

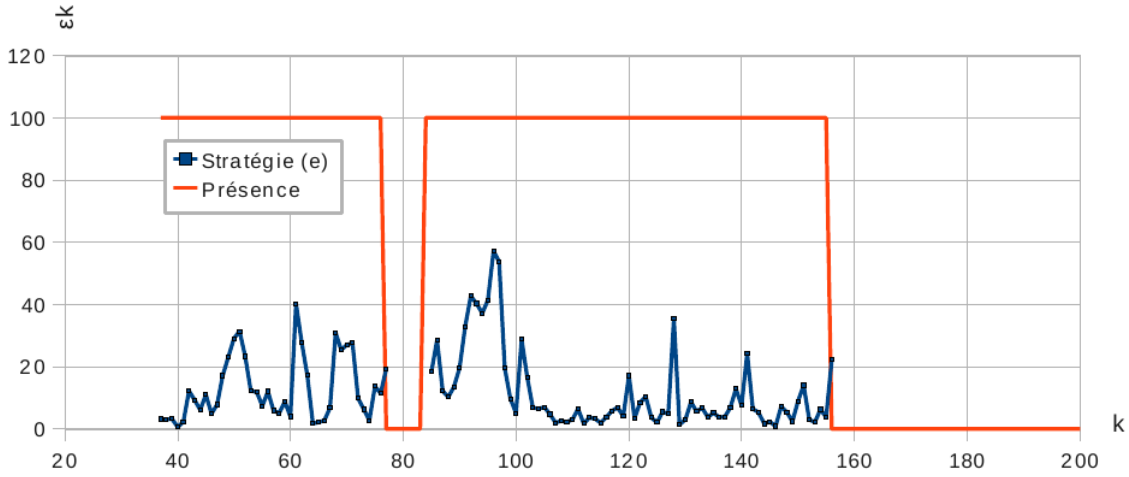
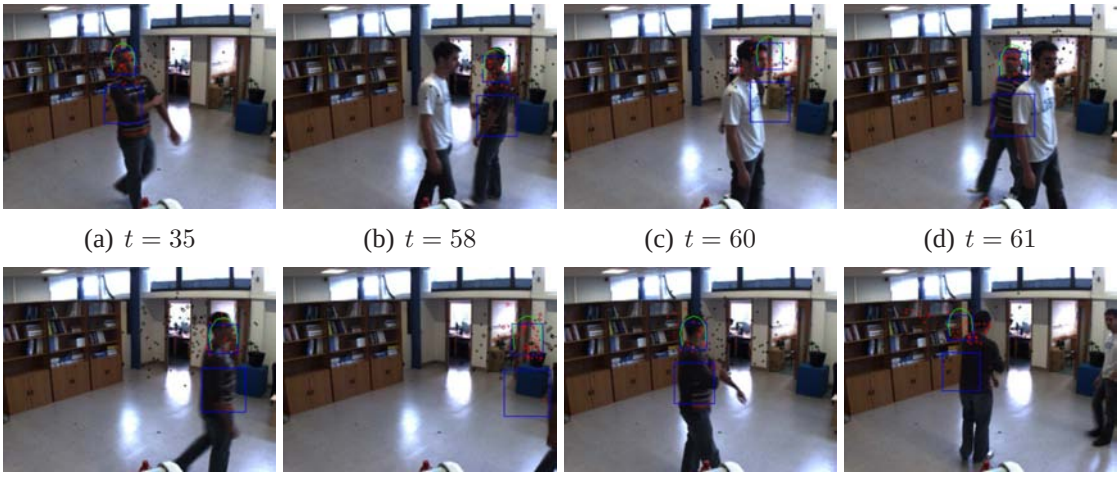


FIG. 3.5 – Evolution de la distance moyenne entre l'estimé MMSE de la stratégie présentée table 3.2(e) et la vérité-terrain pour une séquence type.



(a) $t = 35$ (b) $t = 58$ (c) $t = 60$ (d) $t = 61$

(e) $t = 73$ (f) $t = 76$ (g) $t = 87$ (h) $t = 90$

FIG. 3.6 – Images clés de la séquence de test #2 utilisée pour l'évaluation de la robustesse du suivi multimodal aux artefacts de l'environnement.

Au temps $t = 9$, l'initialisation automatique du filtre est basée sur la correspondance entres l'identification visuelle et RF. Entre $t = 33$ et $t = 39$, la personne cible est occultée par un passant (en pull orange). La fusion de données au niveau de la fonction d'importance et de la fonction de mesure permet alors de rester accroché sur la bonne cible. Il en est de même pour de plus longues occultations entre $t = 92$ et $t = 102$ et entre $t = 222$ et $t = 227$. La séquence complète est disponible à l'adresse `homepages.laas.fr/tgerma/these`.

De plus, on peut observer que l'échantillonnage suivant la fonction d'importance permet de

FIG. 3.7 – Images clés de la séquence de test utilisée pour les évaluations. Chaque ligne donne l'image d'origine (a), la carte de saillance associée (b), les particules après l'étape de pondération (étape 7 de l'algorithme 3.1) (c) et les particules après l'étape de rééchantillonnage (étape 11 de l'algorithme 3.1) + l'estimé MMSE (d).

positionner les particules sur les zones pertinentes de l'espace d'état avant pondération, alors que l'étape de rééchantillonnage permet de reconcentrer le nuage autour de la cible. Ces deux étapes sont donc nécessaires et complémentaires.

3.7 Conclusion et perspectives

Ce chapitre a présenté notre stratégie de fusion de données hétérogènes au sein d'un filtre particulaire pour le suivi de personne en environnement humain. Ces développements apportent un niveau supplémentaire à notre système global d'interaction Homme / Robot et permettent au robot de percevoir l'utilisateur quelles que soient les conditions d'illumination ou la situation Homme / Robot.

La contribution principale de ce chapitre concerne le développement d'un algorithme de suivi multimodal de personnes combinant la richesse de l'information visuelle et l'identification du capteur RF. Notre stratégie utilise l'algorithme de ICONDENSATION, une proposition d'échantillonnage conduite par des données hétérogènes ainsi qu'un algorithme d'échantillonnage par rejet. Ici, la fusion de données s'effectue principalement au niveau de la fonction d'importance grâce à l'utilisation de cartes de saillance alors que la majeure partie des approches de la littérature se focalise sur une fusion de données au sein de la fonction de mesure. A notre connaissance, l'utilisation d'une telle fusion multimodale de données au sein d'un filtre à particules est unique dans la communauté Robotique et/ou Vision. De même, la fusion de données vision et RFID a été peu étudié dans un contexte robotique mobile autonome.

Nous avons évalué notre fonction d'importance par rapport à des fonctions d'importance plus classiques. Il s'avére que l'identification de visages et l'utilisation du badge RFID améliorent la robustesse de notre stratégie en présence de plusieurs personnes dans le champ de vue du robot, là où d'autres méthodes plus conventionnelles pêchent par une faible capacité de réinitialisation. Les évaluations hors-lignes sur des séquences aquises depuis notre robot mobile montrent que notre système bénéficie de capacités permettant : (1) de rester accroché sur la personne cible dans un environnement encombré et en perpétuel changement, (2) de retrouver automatiquement le contact visuel avec la personne cible après une complète occultation voir une disparition temporaire du champ de vue. Cet algorithme a ensuite été testé en conditions réelles afin d'en évaluer ses performances en situations contraintes par notre contexte applicatif. Notre algorithme reste alors robuste aux différents artefacts de l'environnement.

Plusieurs travaux et investigations sont en cours concernant le suivi de personne et la fusion de données hétérogènes.

Tout d'abord, il serait cohérent d'utiliser la complémentarité des différentes données pour détecter les échecs/décrochages du filtre. Une telle fonctionnalité pourrait permettre d'adapter la stratégie d'échantillonnage en conséquence afin de rétablir le contact entre le robot et la cible plus efficacement et plus rapidement. Des études préliminaires, basés sur les travaux de Azimi-Sadjadi *et al.* [Azimi-Sadjadi and Krishnaprasad, 2004], sont actuellement menés. Ces études reposent sur la mise en concurrence de deux filtres à particules ayant chacun une dynamique différente. Lorsque la cible est suivie, les deux filtres ont un comportement similaire. Lorsque

la cible disparaît ou est perdue, les filtres ont un comportement qui diffère. Ces observations permettent alors de détecter une possible erreur de suivi. Tout l'intérêt d'une telle approche serait de pouvoir détecter en temps réel un échec du filtre qui peut intervenir malgré la stratégie de ré-initialisation mise en place.

Dans un deuxième temps, comme indiqué au cours du chapitre, les mesures permettant d'évaluer le facteur d'échelle s sont peu discriminantes. L'utilisation de mesures de la distance Homme / Robot pourrait être envisagée par la fusion au sein du filtre de mesures issues d'autres capteurs tels que le laser ou l'utilisation de mesures stéréovision éparses. Des travaux sur la fusion de données issues du laser dans la fonction d'importance, mais aussi au sein de la fonction de vraisemblance sont actuellement en cours.

Chapitre 4

Suivi multi-personnes pour la détection d'obstacles

Lors d'une mission d'interaction où le robot accompagne son utilisateur dans un espace public, le robot doit être capable de percevoir les autres individus présents dans son voisinage afin, le cas échéant, de pouvoir les éviter. En effet, un robot de service agissant en environnement humain hautement dynamique doit interagir avec son interlocuteur tout en gérant les occultations des passants. Dans la relation établie lors d'une interaction Homme / Robot, ces derniers peuvent alors être perçus comme des obstacles mobiles. Il est donc nécessaire de connaître les déplacements des personnes situées au voisinage immédiat du robot afin d'effectuer des mouvements coordonnés, dans l'intérêt d'une interaction sociale – et sociable – complète *i.e.* afin de commander le déplacement du robot vers la personne cible tout en évitant les passants, leur cédant le passage lorsqu'ils s'approchent du robot.

Une telle fonctionnalité implique le suivi conjoint des différentes personnes présentes autour du robot lors de ses déplacements, ce qui peut entraîner une augmentation considérable des temps de calcul et, par voie de conséquence, induire une perte de réactivité de la part du robot. Il est donc nécessaire de définir une stratégie de suivi adaptée à un nombre variable de cibles. L'interaction qui en découle est alors qualifiée de passive car le robot conserve son attention sur l'utilisateur principal. Seuls les passants susceptibles d'induire une modification des déplacements du robot, *i.e.* au voisinage immédiat du robot, sont perçus.

Ce chapitre présente une méthode de suivi multi-cibles utilisant des données issues de différents capteurs afin de caractériser les déplacements des personnes situées aux alentours du robot. En effet, bien que donnant de bons résultats au regard de la relation exclusive utilisateur - robot, le filtre particulaire décrit au chapitre précédent permet difficilement de gérer un nombre important et variable de cibles en même temps. L'apparition et la disparition de cibles au sein même du vecteur d'état entraîne des sauts de la dimension du vecteur. Au delà (1) des problèmes d'associations de données inhérents au contexte applicatif, (2) des problèmes d'occultations de cibles, la méthode proposée doit être capable de gérer plusieurs cibles en même temps, mais aussi l'apparition d'une nouvelle cible au voisinage du robot ou la disparition d'une autre. De plus, il est intéressant de ne pas utiliser simplement les détections à chaque instant,

mais de permettre une analyse spatio-temporelle de chaque cible afin de prendre en compte leur dynamique lors de la phase d'évitement, pour faciliter leurs déplacements conjoints et de filtrer les différentes détections. Plusieurs stratégies de filtrage sont présentées dans la littérature permettant de gérer ces évènements. Parmis les plus récentes, il existe des stratégies de type filtre particulaire MCMC tout à fait adaptées à ces problèmes de sauts du vecteur d'état. Ce chapitre se concentre sur l'implémentation d'une stratégie de suivi multi-cibles basée sur les fonctions de détections et d'identifications, vues notamment au chapitre 2, et sur une fonction de proposition multi-données utilisant les cartes de saillance introduite au chapitre 3, ainsi que sur d'autres détecteurs complémentaires *i.e.* la détection visuelle de personnes et la détection laser. Nous proposons donc d'étendre notre technique de fusion de données hétérogènes à une problématique de suivi multi-personnes.

Ce chapitre, résultant des travaux de fin de thèse, est structuré comme suit. La section 4.1 présente un état de l'art des méthodes de suivi multi-cibles utilisées dans le cas de suivi de personnes. La section 4.2 détaille notre approche basée sur un filtre particulaire et une méthode d'échantillonnage MCMC dont le formalisme est introduit en section 4.3. La section 4.4 présente notre stratégie de suivi multi-cibles et détaille l'implémentation de cette dernière dans notre contexte applicatif. Des évaluations qualitatives et quantitatives préliminaires sont présentées et commentées dans la section 4.5. La section 4.6 résume nos contributions, les investigations en cours ainsi que les perspectives sur ces travaux. L'évaluaton de la tâche robotique relative à ces travaux sera traitée dans le chapitre 5.

4.1 Etat de l'art

Les techniques de suivi visuel multi-personnes ont été largement abordées ces dernières années au sein de la communauté Vision du fait de leur facilité d'application à l'ensemble des problématiques de vidéosurveillance *e.g.* dans les lieux publics ou privatifs (cf. survey de [Gabriel et al., 2003]), mais aussi dans un cadre plus restreint depuis une plateforme mobile. L'un des principaux challenges des approches de suivi multi-cibles est d'estimer simultanément les déplacements des personnes dans la scène observée, celles-ci pouvant *a priori* entrer ou sortir de la scène, s'approcher les unes des autres. Les objectifs sont alors (i) de détecter correctement les entrées, sorties et occultations temporaires des cibles dans l'espace observé *i.e.* caractériser le statut de chaque cible, (ii) de suivre au cours du temps chaque cible dans le plan image ou le plan du sol.

Au cours des dernières années, les techniques séquentielles de Monte Carlo, notamment le filtre particulaire présenté au chapitre 3, ont été fréquemment utilisées pour le suivi multi-cibles [Isard and MacCormick, 2001; Cho et al., 2007; Qu et al., 2007]. Le filtre particulaire est particulièrement adapté à la fusion de données issues de capteurs hétérogènes. Les techniques de filtrage particulaire appliquées au suivi multi-cibles reposent sur deux stratégies possibles :

– une solution décentralisée basée sur des filtres déportés *i.e.* un filtre par cible [Ryu and Huber, 2007; Qu et al., 2007],

– une solution centralisée utilisant un espace d'état variable mais commun *i.e.* un seul filtre avec un vecteur d'état concaténant l'ensemble des cibles [Isard and MacCormick, 2001; Khan et al., 2005; Smith et al., 2005].

En ce qui concerne les approches décentralisées, visant à initialiser un filtre par cible suivie [Ryu and Huber, 2007], bien que très performantes lorsqu'il s'agit de suivre un nombre précis de cibles [Qu et al., 2007], elles s'avèrent peu adaptées (i) lorsqu'un grand nombre de cibles (*i.e.* au delà de trois) sont présentes en même temps dans la scène, (ii) à un nombre variable de cibles, (iii) à la gestion de l'identifiant de la cible. De plus, l'utilisation de filtres distribués soulève de problème de l'interaction entre les différents filtres et donc de l'association des données. En effet, la modélisation des intéractions entre les cibles rend le processus lourd et combinatoire au fur et à mesure que le nombre de cible croît. Qu *et al.* [Qu et al., 2007] proposent d'utiliser des filtres distribués interactifs en implémentant une stratégie de "répulsion magnétique" entre les différentes cibles lorsque ces dernières tendent à se rapprocher. Cependant, une telle stratégie ne permet pas de gérer un grand nombre de filtres ou un nombre variable de filtres en parallèle car elle s'avère extrêmement coûteuse en temps de calcul. De plus, dans notre contexte, l'espace observé est plutôt restreint avec des cibles par définition en interaction. Il est donc difficile de considérer une approche de suivi multi-cibles décentralisée dans notre contexte impliquant de nombreuses apparitions et disparitions de cibles au cours du temps. Rappelons que nous nous situons dans un contexte d'application nécessitant une grande réactivité de la part du robot. Il n'est donc pas concevable d'utiliser une approche trop dépendante du nombre de cibles *i.e.* qui instancie un nouveau filtre pour chaque cible.

A contrario, les approches de suivi multi-cibles centralisées sont très répandues. Leur intérêt vient du fait qu'elles ne gèrent qu'un seul vecteur d'état regroupant l'ensemble des cibles *i.e.* un seul filtre pour gérer l'ensemble des cibles autour du robot. Cependant, bien qu'ils permettent une gestion implicite du problème d'association de données, les filtres particulaires seuls souffrent d'un manque de performances lorsqu'il s'agit d'estimer un état de trop grande dimension [Kurazume et al., 2008]. Ceci mène souvent à la fusion ou à la perte de cibles au détriment d'une unique.

Une alternative vise à utiliser les méthodes MCMC (pour *Markov Chain Monte Carlo*) [Zhao and Nevatia, 2004; Khan et al., 2005; Smith et al., 2005; Kurazume et al., 2008; Yao and Odobez, 2008]. En effet, les MCMC permettent de gérer un vecteur d'état de dimension variable (*i.e.* les positions des cibles) mais aussi les sauts dans les composantes du vecteur d'état (*i.e.* le nombre de cibles, leur état). Dans [Khan et al., 2005], Khan *et al.* utilisent un filtre particulaire basé sur les MCMC pour suivre les mouvements de fourmis dans une image. Bien qu'une hypothèse forte soit faite sur les détections, les résultats obtenus permettent de confirmer l'apport des méthodes basées sur les MCMC pour le suivi multi-cibles sur une stratégie de filtrage particulaire. D'autres études sont faites dans le domaine de la vidéosurveillance où là encore, les MCMC sont associés aux filtres particulaires pour suivre les nombreuses personnes présentes dans la scène [Zhao and Nevatia, 2004; Smith et al., 2005; Bardet and Chateau, 2008; Yao and Odobez, 2008]. Ces différents travaux utilisent une soustraction de fond pour détecter la présence de cibles ainsi que leur apparition/disparition. Ainsi, Khan *et al.* [Khan et al., 2005] ont démontré la supériorité des

méthodes de filtrage particulaire MCMC par rapport au filtrage particulaire par importance dans le contexte du suivi multi-cibles.

Bien que ces études aient démontré l'attrait des méthodes MCMC couplées à un filtre particulaire pour le suivi multi-cibles, très peu d'entre elles sont utilisées dans le cadre d'une plateforme mobile, ou utilisent la fusion de données multi-capteurs, notamment dans la fonction de proposition. A notre connaissance, seuls Kurazume *et al.* dans [Kurazume et al., 2008] couplent un filtre particulaire et un MCMC pour suivre plusieurs personnes à l'aide de caméras et de lasers déportés, instrumentant un environnement intérieur. Cependant, ici encore, seules les cibles sont mobiles alors que notre système doit gérer les mouvements conjoints des capteurs et des cibles.

Au delà des applications de vidéosurveillance, certains travaux se concentrent sur le suivi de cibles multiples et mobiles depuis une caméra mobile [Chen and Chan, 2007; Ess et al., 2008]. Cependant, le suivi multi-cibles depuis une plateforme mobile autonome reste un vrai challenge du fait de l'environnement hautement dynamique et des ressources calculatoires limitées. Chen *et al.* [Chen and Chan, 2007] utilisent les équations du flot optique dans une séquence d'image pour isoler les différents corps en mouvement. Chaque point d'intérêt est alors associé à une cible avec une certaine probabilité, mise à jour au cours du temps. Cette heuristique permet alors de gérer plus ou moins efficacement les occultations. De même, Ess *et al.* [Ess et al., 2008] utilisent l'odométrie visuelle afin de segmenter les personnes en mouvements dans une scène acquise depuis une plateforme mobile. Un filtre de Kalman prédit alors la position de la caméra afin de faciliter les détections des passants. Cependant, cette problématique reste peu abordée dans la littérature et peu applicable dans le cadre d'un suivi multi-cibles impliquant des contraintes temporelles fortes compatibles avec notre contexte applicatif.

4.2 Notre approche

A l'instar de [Khan et al., 2005; Smith et al., 2005], notre approche a pour but de combiner les avantages : (i) d'un filtre particulaire centralisé, et (ii) d'un échantillonnage par MCMC. Nous proposons donc une stratégie d'échantillonnage des particules par MCMC, tirant parti de données multisensorielles. Les paramètres de chaque cible sont sujets à un échantillonnage par la fonction de proposition similaire à celle décrite en section 3.4.

Notre objectif est ici de suivre un ensemble de personnes cibles présentes autour du robot tout au long d'une mission. Pour cela, nous cherchons à estimer pour chaque cible, ses coordonnées (x_k, y_k) dans le repère du robot *i.e.* le plan du sol. Le vecteur d'état $\mathbf{X}_k$ regroupe alors l'ensemble des cibles $\{\mathbf{X}_{k,j} = (x_{k,j}, y_{k,j})\}_{j \in \mathbf{I}_k}$, défini sur un espace d'état de dimension variable $\chi = \bigcup_{p=1}^{P} \mathbb{R}^{pN_p}$ avec N_p le nombre de paramètres à estimer par cible (ici deux) et où $\mathbf{I}_k$ est l'ensemble de taille P (*i.e.* $card(\mathbf{I}_k) = P$) des identifiants associés à chaque cible.

Dans notre contexte, le nombre de cibles présentes au voisinage du robot ($[0; 5]$m) peut être très variable d'un instant à l'autre. Chaque particule de notre filtre est donc décrite dans un espace d'état de dimension variable. L'échantillonnage de la distribution à estimer *a posteriori* $p(\mathbf{X}_k|z_{1:k})$ est alors dirigé par un algorithme $\mathrm{RJ-MCMC}$ (pour *Reversible Jump* MCMC) qui se veut très efficace dans de tels cas [Khan et al., 2005; Smith et al., 2005]. L'algorithme de

filtrage particulaire basé $\mathrm{RJ-MCMC}$ permet alors aux cibles d'entrer et sortir de la scène. En effet, l'échantillonnage par $\mathrm{RJ-MCMC}$ permet au vecteur d'état de "sauter" d'un sous-espace d'état vers un autre sous-espace d'état de dimension supérieure dans le cas de l'entrée d'une cible ou inférieure dans le cas d'une sortie. Dans cette approche, une fusion de données multi-sensorielle est aussi considérée afin de bénéficier des avantages de chaque capteur mis en jeu (vision, laser, RFID). De plus, le choix d'un filtrage particulaire basé $\mathrm{RJ-MCMC}$ se justifie par le fait que les capteurs embarqués ont un champ de vue limité et sont mobiles, ce qui induit un grand nombre d'entrées / sorties à gérer de manière efficace.

4.3 Généralités sur le filtre particulaire MCMC pour le suivi multi-cibles

Les filtres particulaires tels que décrits dans le chapitre 3 peuvent être étendus afin de gérer plusieurs cibles. Pour ce faire, les approches de filtrage particulaire basées MCMC proposent de remplacer l'étape d'échantillonnage par importance de l'algorithme classique SIR par une échantillonneur MCMC. En effet, l'échantillonnage par importance des particules, tel que présenté en section 3.4, est peu efficace lorsqu'il s'agit d'espaces d'état de grande dimension. Pour palier à ce problème, de nombreux travaux utilisent avec succès des méthodes MCMC [Zhao and Nevatia, 2004; Khan et al., 2005].

Initialement, les modèles à espace d'état joints ont été proposés dans le but de suivre efficacement plusieurs cibles qui interagissent entre elles au moyen d'une stratégie hybride combinant un filtre particulaire et une étape d'échantillonnage basée sur un MCMC. Cette idée, introduite dans [Khan et al., 2005; Smith et al., 2005], permet de gérer un nombre de cibles variables par l'utilisation d'une technique MCMC trans-dimensionnelle à la fois pour les variables continues (position des cibles) et dicrètes (nombre et états des cibles). Ceci permet alors la génération d'une approximation de la distribution *a posteriori* de l'état à estimer par un ensemble d'échantillons non pondérés définis sur un espace d'état de dimension variable.

Les méthodes MCMC sont des méthodes d'échantillonnage à partir de distribution de probabilité. Ces dernières se basent sur la construction de chaînes de Markov qui ont pour lois stationnaires les distributions à échantillonner. Certaines utilisent une dynamique de marche aléatoire *i.e.* Metropolis-Hastings ou Gibbs [Spall, 2002; Mori and Chong, 2008], alors que d'autres algorithmes, plus complexes, introduisent des contraintes sur les parcours afin d'accélérer la convergence *i.e.* Monte Carlo ou Surrelaxation successive. L'état de la chaîne de Markov après un grand nombre d'itérations est alors considéré comme une approximation de la distribution de probabilité à estimer. La qualité de l'approximation est améliorée proportionnellement au nombre d'itérations initiales appelées *Burn in*.

La principale difficulté est de déterminer le nombre d'étapes nécessaires à la convergence vers une distribution stationnaire avec une erreur acceptable *i.e.* définir le nombre d'itérations du *Burn in*. Les méthodes MCMC sont utilisées dans le but de diversifier les échantillons au sein d'un filtre particulaire.

A la différence du filtre particulaire décrit précédemment, la densité de probabilité *a posteriori* de l'état $p(\mathbf{X}_k|z_{1:k})$ est approximée par un ensemble de particules non pondérées $\{\mathbf{X}_k^{(i)}\}_{i=1}^{N_k}$. Une estimée de la distribution $p(\mathbf{X}_k|z_{1:k})$ est alors obtenue suivant le vecteur d'état "le plus représenté" au sein du nuage de particules $\{\mathbf{X}_k^{(i)}\}_{i=1}^{N_k}$.

4.3.1 Algorithme générique d'un filtre particulaire MCMC

La stratégie de filtrage particulaire MCMC consiste à estimer récursivement la densité de probabilité de l'état représentant la configuration jointe multi-objet $\mathbf{X}_k$ à l'instant image k. Au delà des paramètres continus à estimer, à chaque cible est associé un identifiant $\mathbf{I}_{k,j}$ permettant de discriminer chaque cible.

La stratégie de filtrage particulaire basée MCMC, présentée par l'algorithme 4.1, permet de simuler une telle chaîne de Markov. A chaque instant k, connaissant la mesure z_k et la description particulaire $\{\mathbf{X}_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}$, N_k particules $\{\mathbf{X}_k^{(i)}\}_{i=1}^{N_k}$ sont générés afin de représenter au mieux la densité de probabilité associée à l'ensemble des cibles. Pour celà, des cibles j sont choisies aléatoirement dans la chaîne générée à l'itération précédente $\mathbf{X}_k^{(n-1)}$ (étape 3) et échantillonnée suivant une fonction de proposition $q(\mathbf{X}_{k,j}^{(n)}|\mathbf{X}_{k-1,j}^{(i)}, z_k)$ (étape 5). Un ratio d'acceptation a entre la particule avant et après dispersion d'une cible est alors calculé (étape 6) afin d'évaluer la pertinence du changement selon une fonction de vraisemblance $p(z_k|\mathbf{X}_k^{(n)})$ et la distribution $q(.)$. Si le changement ne correspond pas (1) à la dynamique propre à la cible, (2) aux mesures courantes, (3) aux contraintes d'interaction entre les cibles, le changement effectué sur l'état initial est rejeté (étape 7).

L'usage courant veut qu'un certain nombre d'itérations N_B (aussi appelé *Burn in*) soit appliqué afin de permettre à l'algorithme de converger vers la distribution stationnaire et de s'affranchir d'une trop forte dépendance avec l'état initial. De plus, dans le but de réduire la correlation entre les échantillons, seul un certain nombre d'échantillons générés par l'algorithme, prélevés à intervalles régulier M (*i.e.* lorsque $(it - N_B) \equiv 0_{[M]}$), est conservé lors d'une étape d'affinage (ou *Thin out*).

Dans le cas de l'algorithme 4.1, la fonction de proposition $q(\mathbf{X}_{k,j}|\mathbf{X}_{k-1,j}, z_k)$ est une fonction dite mono-cible identique à celle définie par l'équation 3.6 *i.e.* elle n'agit que sur une seule cible $\mathbf{X}_{k,j}$ de l'échantillon $\mathbf{X}_k$ et elle est guidée en partie par z_k. En effet, une cible j de l'échantillon $\mathbf{X}_k$ est sélectionnée aléatoirement et cette dernière est soumise à une dynamique définie par l'état précédent $\mathbf{X}_{k-1}$ et les mesures z_k à l'instant k.

4.3.2 Extension au filtre particulaire $\mathrm{RJ-MCMC}$

Bien que donnant de très bons résultats dans le cadre de l'échantillonnage d'un nombre donné de cibles, l'algorithme MCMC à lui tout seul n'autorise pas l'apparition ou la disparition d'une ou plusieurs cibles *i.e.* il n'est pas adapté à un nombre variable de cibles.

Afin de gérer un nombre variable de cibles de manière automatique, l'étape d'échantillonnage MCMC (algorithme 4.1) peut alors être remplacée par un échantillonnage $\mathrm{RJ-MCMC}$ (pour

ALG. **4.1** Algorithme générique de filtrage particulaire MCMC pour le suivi d'un nombre fixe de cibles.

ENTRÉES: $[\{\mathbf{X}_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}, z_k]$

SORTIES: $[\{\mathbf{X}_k^{(i)}\}_{i=1}^{N_k}]$

1: Générer une particule initiale $\mathbf{X}_k^0$ telle que

$$\mathbf{X}_{k,j}^0 \sim p(\mathbf{X}_{k,j}|\mathbf{X}_{k-1,j}^{(i)}) \,, \forall \mathbf{X}_{k-1,j}^{(i)} \in \mathbf{X}_{k-1}^{(i)}$$

où i est l'indice d'une particule choisie aléatoirement dans $\{\mathbf{X}_{k-1}^{(i)}\}$, $n = 0$

2: **pour** $it = 0, \ldots, N_B + MN_k$ **faire**

3: Choisir aléatoirement une cible j de $\mathbf{X}_k^{(n)}$

4: Choisir aléatoirement une particule i de $\{\mathbf{X}_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}$ telle que $\mathbf{I}_{k,j}^{(n)} \in \{\mathbf{I}_{k-1}^{(i)}\}$

5: Proposer un nouveau vecteur d'état $\mathbf{X}_k^{(n)'}$ tel que $\mathbf{X}_{k,j}^{(n)'} \sim q(\mathbf{X}_{k,j}^{(n)}|\mathbf{X}_{k-1,j}^{(i)}, z_k)$

6: Calculer le ratio d'acceptation

$$a = \frac{p(z_k|\mathbf{X}_k^{(n)'})}{p(z_k|\mathbf{X}_k^{(n)})} \frac{q(\mathbf{X}_{k,j}^{(n)}|\mathbf{X}_{k,j}^{(n)'}, z_k)}{q(\mathbf{X}_{k,j}^{(n)'}|\mathbf{X}_{k,j}^{(n)}, z_k)}$$

7: Accepter $\mathbf{X}_k^{(n)'}$ avec une probabilité a *i.e.* $\mathbf{X}_k^{(n)} = \mathbf{X}_k^{(n)'}$, sinon, rejeter $\mathbf{X}_k^{(n)'}$

—Sélection de la particule après les étapes de "Burn in" et "Thin out"—

8: **si** $(it - N_B) \equiv 0_{[M]}$ **alors**

9: $\{\mathbf{X}_k\} = \{\mathbf{X}_k\} \cup \mathbf{X}_k^{(n)}$, $n = n + 1$

10: **fin si**

11: **fin pour**

Reversible Jump MCMC), une généralisation du MCMC aux espaces d'état de dimension variable [Waagepetersen and Sorensen, 2001; Khan et al., 2005]. L'algorithme $\mathrm{RJ-MCMC}$, présenté par l'algorithme 4.2, estime alors les configurations jointes $\mathbf{X}_k \in \chi$. Le principe est ensuite de définir un ensemble fini d'"évènements" $\{\mathbf{m}\}$ qui peuvent (i) augmenter la taille de l'espace d'état *i.e.* apparition d'une cible, (ii) diminuer la taille de l'espace d'état *i.e.* disparition d'une cible, (iii) laisser le vecteur d'état dans le même sous-espace. Un évènement permettant de changer la dimension du vecteur d'état est appelé un saut (ou *jump*). Un nouvel état $\mathbf{X}_k^{'}$ est alors proposé par la fonction de proposition $q_{\mathbf{m}}(\mathbf{X}_k^{'}|\mathbf{X}_{k-1}, z_k)$ relative au saut $\mathbf{m}$, où $\mathbf{m}$ est déterminé par $q(\mathbf{m})$, une fonction de proposition du saut $\mathbf{m}$ relative au contexte et définie empiriquement. Chaque saut $\mathbf{s} \in \{\mathbf{m}\}$ doit avoir son saut inverse $\mathbf{s}^{'} \in \{\mathbf{m}\}$ *e.g* à chaque saut symbolisant l'apparition d'une cible doit correspondre un saut inverse représentant la disparition d'une cible. Cette condition permet d'éviter le "confinement" du vecteur d'état à un optimum local de l'espace d'état. L'algorithme d'échantillonnage $\mathrm{RJ-MCMC}$ permet donc de gérer dynamiquement l'apparition et la disparition de cibles au sein même du vecteur d'état.

La méthode de filtrage particulaire par $\mathrm{RJ-MCMC}$ présentée par l'algorithme 4.2 est plus

ALG. **4.2** Algorithme de filtrage particulaire $\mathrm{RJ-MCMC}$ pour le suivi d'un nombre variable de cibles.

ENTRÉES: $[\{\mathbf{X}_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}, z_k]$

SORTIES: $[\{\mathbf{X}_k^{(i)}\}_{i=1}^{N_k}]$

1: Générer une particule initiale $\mathbf{X}_k^0$ telle que

$$\mathbf{X}_{k,j}^0 \sim p(\mathbf{X}_{k,j}|\mathbf{X}_{k-1,j}^{(i)}) \,, \forall \mathbf{X}_{k-1,j}^{(i)} \in \mathbf{X}_{k-1}^{(i)}$$

où i est l'indice d'une particule choisie aléatoirement dans $\{\mathbf{X}_{k-1}^{(i)}\}$, $n = 0$

2: **pour** $it = 0, \ldots, N_B + MN_k$ **faire**

3: Choisir un saut $\mathbf{m} \sim q(\mathbf{m})$

4: Choisir une cible j de $\mathbf{X}_k^{(n)}$

5: Choisir aléatoirement une particule i de $\{\mathbf{X}_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}$ telle que $\mathbf{I}_{k,j}^{(n)} \in \{\mathbf{I}_{k-1}^{(i)}\}$

6: Proposer un nouveau vecteur d'état $\mathbf{X}_k^{(n)'}$ tel que :

$$\mathbf{X}_{k,j}^{(n)'} \sim q_{\mathbf{m}}(\mathbf{X}_{k,j}^{(n)}|\mathbf{X}_{k-1,j}^{(i)}, z_k) \tag{4.1}$$

7: Calculer le ratio d'acceptation

$$a = \frac{p(z_k|\mathbf{X}_k^{(n)'})}{p(z_k|\mathbf{X}_k^{(n)})} \frac{q(\mathbf{m}')}{q(\mathbf{m})} \frac{q_{\mathbf{m}'}(\mathbf{X}_{k,j}^{(n)}|\mathbf{X}_{k,j}^{(n)'}, z_k)}{q_{\mathbf{m}}(\mathbf{X}_{k,j}^{(n)'}|\mathbf{X}_{k,j}^{(n)}, z_k)} \tag{4.2}$$

8: Accepter $\mathbf{X}_k^{(n)'}$ avec une probabilité a *i.e.* $\mathbf{X}_k^{(n)} = \mathbf{X}_k^{(n)'}$, sinon, rejeter $\mathbf{X}_k^{(n)'}$

—Sélection de la particule après les étapes de "Burn in" et "Thin out"—

9: **si** $(it - N_B) \equiv 0_{[M]}$ **alors**

10: $\{\mathbf{X}_k\} = \{\mathbf{X}_k\} \cup \mathbf{X}_k^{(n)}$, $n = n + 1$

11: **fin si**

12: **fin pour**

adapté au contexte applicatif impliquant un nombre variable de cibles. En effet, sur la base de l'échantillonnage MCMC, il permet de générer un ensemble de N_k particules $\{\mathbf{X}_k^{(i)}\}_{i=1}^{N_k}$ de dimension variable candidates à l'instant k à partir d'un jeu de N_{k-1} particules $\{\mathbf{X}_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}$ à l'instant $k-1$. La principale différence est donc l'ajout de sauts permettant au vecteur d'état de représenter la configuration jointe d'un nombre variable de cibles.

4.4 Implémentation du suivi multi-cibles

Un problème lors du suivi multi-cibles est l'association des données. En effet, la majeure partie des approches supposent que chaque cible détectée soit identifiée de manière sûre et unique. Or, dans notre contexte, et au regard des différentes méthodes de détection et d'identification

de personnes présentées précédemment *i.e.* détection de couleur peau et laser, identification visuelle et RF de personnes, il est difficile d'associer avec certitude les informations fournies par les différents capteurs. En effet, souvent, plusieurs détections de visages ou de zones de couleur peau correspondent à une seule détection de badge RFID. Nous proposons donc une méthode, dans la veine de la fonction d'importance présentée au chapitre 3, permettant ici encore de définir une fonction de proposition très discriminante par la fusion de données issues de différents capteurs et combinant des données laser, vision et RFID. A l'instar des résultats obtenus au chapitre 3, on peut supposer que si notre fonction de proposition est assez discriminante, alors, l'échantillonnage le sera aussi entraînant une diminution du nombre d'itérations nécessaires lors des phases de *burn in* et *thin out* par rapport à une fonction n'utilisant que la dynamique.

Il est alors nécessaire de définir les sauts de notre système de suivi multi-cibles ainsi que les fonctions de propositions et les ratios d'acceptation associés à chaque saut.

Rappelons que nous cherchons à estimer les paramètres (x, y) qui composent le vecteur d'état $\mathbf{X}_{k,j}$ de la cible j à l'instant k. Concernant la dynamique $p(\mathbf{X}_{k,j}|\mathbf{X}_{k-1,j})$, tout comme précédemment, les déplacements d'un humain sont difficiles à caractériser. Cette faible connaissance est représentée par la définition du vecteur d'état $\mathbf{X}_{k,j} = [x_{k,j}, y_{k,j}]'$ et l'évolution de ses paramètres suit un modèle indépendant de marche aléatoire gaussienne $p(\mathbf{X}_{k,j}|\mathbf{X}_{k-1,j}) = \mathcal{N}(\mathbf{X}_{k,j}; \mathbf{X}_{k-1,j}, \Sigma)$ où la covariance $\Sigma = \mathrm{diag}(\sigma_x^2, \sigma_y^2)$.

4.4.1 Pré-requis sur les détections

La fonction de proposition $q_{\mathbf{m}}(.)$ (équation 4.1 de l'algorithme 4.2) s'inscrit dans la veine de la fonction d'importance définie par l'équation 3.6 d'après l'ensemble des détecteurs $\pi(\mathbf{X}_{k,j}|z_k^l)$, la dynamique $p(\mathbf{X}_{k,j}|\mathbf{X}_{k-1,j})$ et une connaissance *a priori* $p_0(\mathbf{X}_{k,j})$.

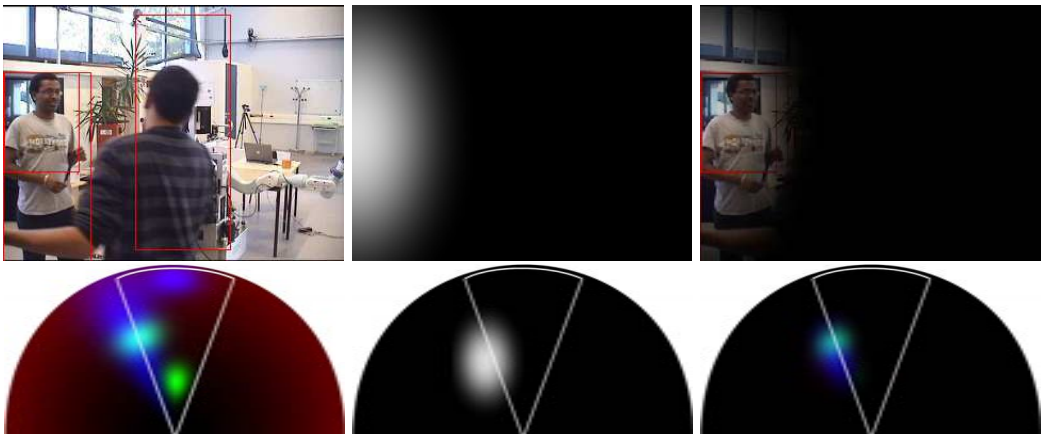
Il est possible d'étendre l'ensemble des détecteurs utilisés au chapitre précédent, *i.e.* les détections de couleur peau $\pi(\mathbf{X}_{k,j}|z_k^c)$ (équation 3.7), l'identification de visages $\pi(\mathbf{X}_{k,j}|z_k^s)$ (équation 3.8) et RFID $\pi(\mathbf{X}_{k,j}|z_k^r)$ (équation 3.9), à l'ensemble des détecteurs définis au chapitre 2, *i.e.* la détection de personnes par vision $\pi(\mathbf{X}_{k,j}|z_k^h)$ et par laser $\pi(\mathbf{X}_{k,j}|z_k^p)$.

La première fonction de proposition $\pi(\mathbf{X}_{k,j}|z_k^p)$ relative aux détections de personnes par laser tel que défini en section 2.3.1 est décrite comme suit. Soit N_p le nombre de personnes détectées par le laser et $\mathcal{P}_i = (\mu_x^{\mathcal{P}_i}, \mu_y^{\mathcal{P}_i})$ la position de la $i^{\text{ème}}$ personne dans le repère du robot. La fonction $\pi(\mathbf{X}_{k,j}|z_k^p)$ est alors décrite par le mélange de gaussiennes suivant :

$$\pi(\mathbf{X}_{k,j}|z^p) = \sum_{j=1}^{N_p} \mathcal{N}(\mathbf{X}_{k,j}; \mathcal{P}_j, \mathrm{diag}(\sigma_x^{\mathcal{P}_j}, \sigma_y^{\mathcal{P}_j})), \tag{4.3}$$

où $(\mu_x^{\mathcal{P}_j}, \sigma_x^{\mathcal{P}_j})$ et $(\mu_y^{\mathcal{P}_j}, \sigma_y^{\mathcal{P}_j})$ sont les valeurs définies dans la section 2.3.1.

De même, la fonction de proposition $\pi(\mathbf{X}_{k,j}|z_k^h)$ relative aux détections visuelles de personnes est basée sur le détecteur défini en section 2.3.2. Soit N_h le nombre de personnes détectées



(a) (b) (c)

FIG. 4.1 – Image masque d'une cible. (a) Détection dans le plan image (haut) et dans le plan du sol (bas)(les détections RFID sont en bleu, les détections laser en vert et la connaissance *a priori* en rouge). (b) Carte de saillance $\mathcal{I}_{k,j}$ de la cible. (c) Application de l'image masque sur les détections. La zone délimitée en blanc correspond au champ de vue de la caméra.

dans le plan image et $\mathbf{p}_i = (u_i, v_i)$ le centre de la région i correspondant à la $i^{\text{ème}}$ personne. La fonction $\pi(\mathbf{X}_{k,j}|z_k^h)$ est décrite en tant que mélange de gaussiennes comme suit :

$$\pi(\mathbf{X}_{k,j}|z^h) = \sum_{j=1}^{N_h} \mathcal{N}(\mathbf{X}_{k,j}^{\mathcal{I}}; \mathbf{p}_j, \text{diag}(\sigma_{u_j}^2, \sigma_{v_j}^2)), \tag{4.4}$$

où σ_{u_j} et σ_{v_j} dépendent respectivement de la largeur et de la hauteur de la personne détectée dans l'image et $\mathbf{X}_{k,j}^{\mathcal{I}}$ correspond à la projection de la cible $\mathbf{X}_{k,j}$ dans le plan image.

L'expression générale de la fonction de proposition $\pi(\mathbf{X}_{k,j}|z_k^l)$ relative à l'ensemble des détecteurs reste alors inchangée par rapport à l'équation 3.5.

4.4.2 Image masque de cible

Afin de guider les évènements de chaque cible, il est nécessaire d'introduire la notion d'image masque relative à une cible définie dans [Bardet and Chateau, 2008]. L'image masque $\mathcal{I}_{k,j}$ d'une cible $\mathbf{X}_{k,j}$ défini une région d'influence de la cible j, *i.e.* tout détecteur $\pi(\mathbf{X}_{k,j}|z_k^l)$ situé dans cette région est associé à la cible j. L'image masque d'une cible j correspond donc à une carte de saillance associée à cette cible. Pour une cible $\mathbf{X}_{k,j}$, l'image masque, définie dans un cadre probabiliste, s'écrit :

$$\forall \mathbf{x}_k = (x_k, y_k) \in \mathcal{I}_{z_k}, \mathcal{I}_{k,j} = \mathcal{N}(\mathbf{x}_k|\mathbf{X}_{k,j}, \Sigma_{\mathbf{X}_{k,j}}), \tag{4.5}$$

où $\Sigma_{\mathbf{X}_{k,j}}$ correspond à la dispersion des cibles $\mathbf{X}_{k-1,j}$ à l'instant $k-1$ et $\mathcal{I}_{z_k}$ est défini dans le repère associé à la mesure z_k *i.e.* dans le plan image pour des détecteurs visuels et dans le plan du sol pour les détecteurs laser et RFID.

L'image masque $\mathcal{I}_k$ correspondant à l'ensemble des cibles d'une particule $\mathbf{X}_k$, et son inverse $\bar{\mathcal{I}_k}$ sont définies par :

$$\mathcal{I}_k = \sum_j \mathcal{I}_{k,j} \text{ et } \bar{\mathcal{I}_k} = 1 - \mathcal{I}_k \tag{4.6}$$

La figure 4.1 illustre d'intérêt de définir une telle carte de saillance associée à chaque cible.

4.4.3 Description des sauts et caractérisation des fonctions de proposition

A partir des différents détecteurs et identificateurs utilisés, il est possible d'obtenir, à chaque instant k, une fonction de proposition $\pi(\mathbf{X}_k|z_k)$ représentant l'ensemble des cibles candidates. De cette connaissance, plus ou moins bruitée suivant les conditions, il est nécessaire de définir les différents sauts (ou évènements) utiles à l'implémentation de l'algorithme d'échantillonnage $\mathrm{RJ-MCMC}$. Sachant que chaque saut $\mathbf{m}$ nécessite la définition de son saut symétrique $\mathbf{m}'$, nous pouvons les classer en trois types :

- mise à jour d'une cible *i.e.* l'évolution de la position d'une cible donnée dans l'espace environnant,
- ajout / suppression d'une cible *i.e.* l'entrée ou la sortie d'une cible dans les alentours du robot,
- permutation de cibles *i.e.* la permutation de leur identifiant.

Nous proposons une fonction de proposition $q_{\mathbf{m}}(.)$ relative à chaque saut $\mathbf{m}$ assurant la sélection (i) d'une cible dans le cas d'une mise à jour $\mathbf{u}$, d'une suppression $\mathbf{s}$ ou d'un ajout $\mathbf{a}$, (ii) d'une paire de cibles dans le cas d'une permutation $\mathbf{p}$.

Ci-après, nous allons donc définir les fonctions de proposition $q_{\mathbf{m}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k)$ ainsi que les ratios d'acceptation associés à chaque évènement $\mathbf{m} \in \{\mathbf{u}, \mathbf{s}, \mathbf{a}, \mathbf{p}\}$.

Mise à jour d'une cible

La mise à jour d'une cible j au sein d'un vecteur d'état est relative à la position de cette dernière autour du robot. Classiquement, la fonction de proposition $q_{\mathbf{u}}(.)$ relative à cet évènement s'inspire de l'équation 3.6. Ainsi, la fonction de proposition $q_{\mathbf{u}}(.)$ prend en compte les détections extraites des mesures courantes et la dynamique de la cible.

Afin d'éviter les problèmes d'association de données, il est nécessaire de limiter les détections $\pi(\mathbf{X}'_{k,j}|z_k)$ à celles associées à la cible, à savoir celles se trouvant dans l'image masque de la cible définie par sa carte de saillance $\mathcal{I}_{k,j}$, *i.e.* $\pi(\mathbf{X}'_{k,j}|z_k).\mathcal{I}_{k,j}$. En effet, cette étape nécessite uniquement la propagation de la cible en fonction des détections $\pi(\mathbf{X}'_{k,j}|z_k)$ associées à la cible et de la dynamique $p(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j})$. La fonction de proposition associée à la mise à jour d'une cible j s'écrit donc :

$$q_{\mathbf{u}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k) = \alpha\pi(\mathbf{X}'_{k,j}|z_k).\mathcal{I}_{k,j} + (1-\alpha)p(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}). \tag{4.7}$$

De plus, cet évènement est auto-reversible *i.e.* $q_{\mathbf{u}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k) = q_{\mathbf{u}}(\mathbf{X}_{k,j}|\mathbf{X}'_{k,j}, z_k)$. Le calcul du ratio d'acceptation a (équation 4.2) devient :

$$a_{\mathbf{m}=\mathbf{u}} = \frac{p(z_k|\mathbf{X}'_k)}{p(z_k|\mathbf{X}_k)} \tag{4.8}$$

Ajout / Suppression d'une cible

L'ajout ou la suppression d'une cible j au sein d'un vecteur d'état est relatif à la position de cette dernière autour du robot. Il est alors possible de définir une connaissance *a priori* $p_0(\mathbf{X}_{k,j})$ représentant la probabilité qu'a une cible $\mathbf{X}_{k,j}$ d'apparaître ou de disparaître de l'ensemble des cibles du vecteur d'état. Cette fonction $p_0(\mathbf{X}_{k,j})$ représente donc les zones d'entrée et de sortie de la scène. Cette probabilité est alors définie dans le repère du robot par :

$$p_0(\mathbf{X}_{k,j}) = 1 - \mathcal{N}(\mathbf{X}_{k,j}; (0,0), \mathrm{diag}(\sigma_{x_0}^2, \sigma_{y_0}^2)), \tag{4.9}$$

où σ_x et σ_y dépendent respectivement de la distance maximum à laquelle une cible doit être perçue ; dans notre cas d'application $\sigma_x = 4.5$m et $\sigma_y = 4.5$m. En effet, une cible ne sera ajoutée que dans le cas où elle se trouverait dans la zone limite de détection *i.e.* environ 4.5m. De même, une cible ne sera supprimée que si elle se trouve dans cette même zone, même si les détecteurs sont actifs à plus longue distance.

Dans le cas d'un ajout de cible, la fonction de proposition $q_{\mathbf{m}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k)$ est entièrement définie par les détections présentes à l'instant k et par $p_0(.)$. De plus, la loi définissant l'ajout d'une cible peut être guidée par l'inverse de l'image masque de la particule $\bar{\mathcal{I}}_k$, afin d'éliminer les zones correspondant aux cibles déjà présentes dans le vecteur d'état. La fonction de proposition associée à l'ajout d'une cible s'écrit donc :

$$q_{\mathbf{a}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k) = \bar{\mathcal{I}}_k.(\alpha\pi(\mathbf{X}'_{k,j}|z_k) + (1-\alpha)p_0(\mathbf{X}'_{k,j})). \tag{4.10}$$

Cette fonction de proposition ne prend alors en compte que les détections qui n'ont pas encore été associées à une cible.

Dans le cas d'une suppression, la fonction de proposition est uniquement définie par la connaissance *a priori* telle que :

$$q_{\mathbf{s}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k) = p_0(\mathbf{X}'_{k,j}). \tag{4.11}$$

En effet, il est difficilement concevable qu'une cible disparaisse alors qu'elle se trouve au milieu de la zone d'observation. *A contrario*, plus une cible se trouve proche de la zone limite de détection (environ 4.5m), plus sa probabilité de disparaître est grande.

Dans ce cas là, les fonctions de proposition relatives à un ajout et à une suppression ne sont pas symétriques *i.e.* $q_{\mathbf{a}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k) \neq q_{\mathbf{s}}(\mathbf{X}_{k,j}|\mathbf{X}'_{k,j}, z_k)$. Le calcul du ratio d'acceptation a (équation 4.2) est alors décrit par :

$$a_{\mathbf{m}=\mathbf{a}} = a_{\mathbf{m}=\mathbf{s}} = \frac{p(z_k|\mathbf{X}'_k)}{p(z_k|\mathbf{X}_k)} \frac{q(\mathbf{m}')}{q(\mathbf{m})} \frac{q_{\mathbf{m}'}(\mathbf{X}_{k,j}|\mathbf{X}'_{k,j}, z_k)}{q_{\mathbf{m}}(\mathbf{X}'_{k,j}|\mathbf{X}_{k,j}, z_k)} \tag{4.12}$$

Permutation de deux cibles

La permutation de deux cibles j_1 et j_2 au sein d'un vecteur d'état consiste à échanger les identifiants de chacune d'entre elles *i.e.* $\mathbf{I}_{k,j_1} \leftrightarrow \mathbf{I}_{k,j_2}$. De même que pour la mise à jour, l'évènement de permutation est aussi auto-reversible. De plus, aucune valeur à estimer n'est modifiée. Le calcul du ratio d'acceptation a (équation 4.2) devient :

$$a_{\mathbf{m=p}} = \frac{p(z_k|\mathbf{X}'_k)}{p(z_k|\mathbf{X}_k)}. \tag{4.13}$$

4.4.4 Description de la fonction de mesure

A l'instar de la fonction de mesure (équation 3.13) utilisée dans l'algorithme de filtrage particulaire défini en section 3.5, il est nécessaire de définir ici une fonction de vraisemblance $p(z_k|\mathbf{X}_k)$ basé sur des mesures persistantes. Dans notre cas, le laser permet d'obtenir une mesure fiable sur l'ensemble de la zone d'observation. La fonction de mesure globale $p(z_k^p|\mathbf{X}_k)$ considére la distance euclidienne entre les mesures laser et l'état à évaluer. Soit $\{z_{k,j}^p(l)\}$ l'ensemble des n points laser consécutifs les plus proches d'une cible $\mathbf{X}_{k,j}$, la fonction de vraisemblance pour une cible donnée j est décrite comme suit :

$$p(z_k^p|\mathbf{X}_{k,j}) = \prod_{i=1}^{n} \left(1 - \exp\left(-\frac{D(\mathbf{X}_{k,j}, z_{k,j}^p(i))^2}{2\sigma_D^2}\right)\right), \tag{4.14}$$

où $D(\mathbf{X}_{k,j}, z_{k,j}^p(i))$ est la distance euclidienne entre le vecteur d'état $\mathbf{X}_{k,j}$ et la $i^{\text{ème}}$ mesure laser $z_{k,j}^p(i)$ dans le plan du sol. L'expression globale de la fonction de vraisemblance basée sur les mesures laser s'écrit alors :

$$p(z_k^p|\mathbf{X}_k) = \prod_j (1 - p(z_k^p|\mathbf{X}_{k,j})). \tag{4.15}$$

De même que pour notre fonction de vraisemblance définie au chapitre précédent, il serait judicieux d'utiliser les mesures images $p(z_k^s|\mathbf{X}_{k,j})$ (équation 3.10) et $p(z_k^c|\mathbf{X}_{k,j})$ (équation 3.12). Cependant, nous avons fait le choix de ne pas les prendre en compte dans notre fonction de mesure globale, car le champ de vue de la caméra ne couvre pas l'ensemble de la zone d'observation, et il serait donc impossible de suivre les multiples cibles qui ne sont pas présentes dans le plan caméra.

Il est nécessaire de prendre en compte les interactions des différentes cibles au sein même du vecteur d'état. Le modèle d'interaction utilisé ici est assez classique. Il consiste à pénaliser les objets en interaction au moyen d'un MRF (pour *Markov Random Field*) défini sur l'ensemble des cibles du vecteur d'état [Khan et al., 2005]. Ce modèle permet de diminuer la vraisemblance des deux cibles qui tendraient à suivre la même personne, *e.g.* lorsque deux personnes se croisent. Le potentiel d'interaction $\Psi(\mathbf{X}_{k,i}, \mathbf{X}_{k,j})$ de deux cibles $\mathbf{X}_{k,i}$ et $\mathbf{X}_{k,j}$ du même vecteur d'état $\mathbf{X}_k$ est défini comme suit :

$$\Psi(\mathbf{X}_{k,i}, \mathbf{X}_{k,j}) \propto \exp\left(-\frac{1}{2}\rho(S_k^{\mathbf{X}_i}, S_k^{\mathbf{X}_j})\right), \tag{4.16}$$

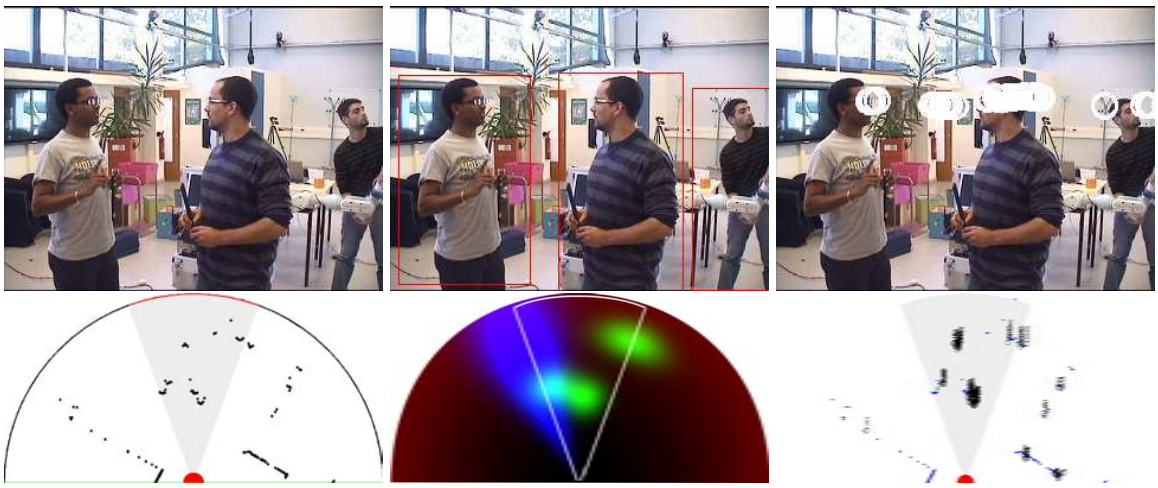
où $S_k^{\mathbf{X}_i}$ et $S_k^{\mathbf{X}_j}$ repésentent les supports respectifs de $\mathbf{X}_{k,i}$ et $\mathbf{X}_{k,j}$, la fonction ρ mesurant les similitudes entre les supports. L'argument de l'exponentielle est alors (i) zéro si les cibles ne sont pas en contact, (ii) minimum si elles correspondent parfaitement.

La fonction de vraisemblance unifiée (étape 7 de l'algorihme 4.2) est donc de la forme :

$$p(z_k|\mathbf{X}_k) = p(z_k^p|\mathbf{X}_k) \prod_{i,j \in \mathbf{I}_k} \Psi(\mathbf{X}_{k,i}, \mathbf{X}_{k,j}). \tag{4.17}$$

La fonction de vraisemblance globale permet de vérifier : (1) la cohérence de l'échantillon vis à vis des mesures courantes (équation 4.15), (2) la cohérence des cibles à l'intérieur de l'échantillon (équation 4.16).

La figure 4.2 montre le fonctionnement de l'échantillonneur MCMC. On peut constater que la majorité des particules est concentrée sur les différentes cibles (figures 4.2(c)) détectées par la caméra, le lecteur RFID, le laser ou plusieurs de ces capteurs combinés (figures 4.2(b)). Il faut remarquer que quelques fausses cibles sont observées dans l'environnement. Elles correspondent à des artefacts et bruits des différents capteurs, *i.e.* des fausses détections, survenus aux itérations précédentes. Cependant, leur influence est négligeable au regard des autres cibles échantillonnées en bien plus grand nombre.



(a) (b) (c)

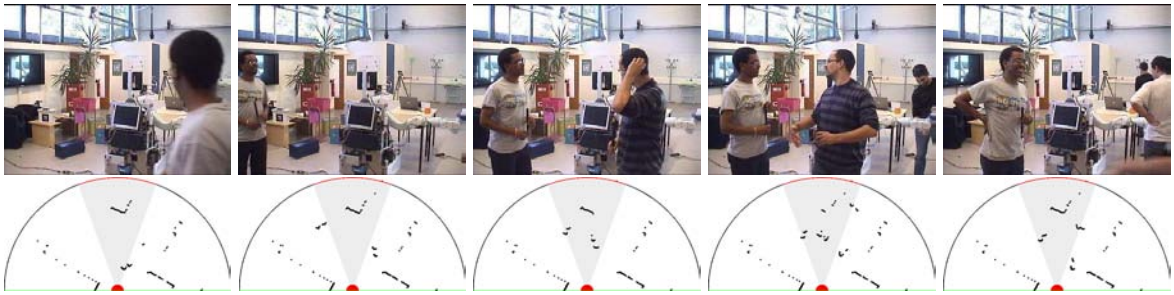
FIG. 4.2 – Détail du fonctionnement de l'échantillonneur MCMC (haut : données caméra, bas : données exprimées dans le repère du robot). (a) Données vision et laser brutes. (b) Détections de personnes et carte de saillance associée. (c) Particules échantillonnées dans le plan du sol suivant l'algorithme 4.2 et leur projection sur le plan image.

Les différents paramètres du filtre (N, $\sigma_{(.)}$, N_B, M) ainsi que des coefficients de pondération ($q(\mathbf{m})$, α, $\kappa_{(.)}$) ont été réglés de manière empirique. Leurs valeurs sont recensées dans la table 4.1.

Symbole	Description	Valeur
$N_k (= N_{k-1})$	nombre de particules du filtre	100
$q(\mathbf{m})$	fonction de proposition des sauts $\mathbf{m} \in \{\mathbf{u}, \mathbf{s}, \mathbf{a}, \mathbf{p}\}$	$(0.70, 0.10, 0.10, 0.10)$
N_B	nombre d'itération de la phase de *Burn in*	$0.2 \times N$
M	nombre d'itération de la phase de *Thin out*	5
(σ_x, σ_y)	écart-type du modèle de marche aléatoire	$(20, 20)$
$(\sigma_{x_0}, \sigma_{y_0})$	écart-type de la connaissance *a priori* (4.9)	$(450, 450)$
α	coeff. des fonctions de proposition $q_{\mathbf{m}}(\mathbf{X}'_{k,j} \vert \mathbf{X}'_{k,j}, z_k)$ (4.7)(4.10)	0.6
$(\kappa_c, \kappa_s, \kappa_r, \kappa_p, \kappa_h)$	coeff. de pondération des détecteurs dans $\pi(x_k^{(i)} \vert z_k^1, \ldots, z_k^L)$	$(0.1, 0.2, 0.2, 0.3, 0.2)$
σ_D	dispersion de la vraisemblance laser $p(z_k^p \vert \mathbf{X}_{k,j})$ (4.14)	50

TAB. 4.1 – Valeurs des paramètres utilisées pour le suivi multi-personnes.

4.5 Evaluations préliminaires et commentaires associés



(a) Séquence #1 impliquant deux personnes, de brèves occultations et un arrière-plan très encombré. Le robot est statique.

(b) Séquence #2 impliquant jusqu'à cinq personnes simultanément, de brèves occultations et un arrière-plan très encombré. Le robot est en mouvement.

FIG. 4.3 – Détails des séquences d'images utilisées pour l'évaluation des performances de notre approche. Chacune des séquences comprend (1) plusieurs personnes, (2) occultations des cibles, (3) apparitions/disparitions du champ de vue, (4) déplacements erratiques des acteurs.

L'algorithme de suivi multi-cible a été prototypé sur un processeur Pentium Dual Core 1.8GHz sous Linux à l'aide de la bibliothèque OpenCV. Les évaluations quantitatives et qualitatives hors-ligne sont ici présentées. La base de données issues de différents capteurs contient plusieurs séquences totalisant plus de 2200 images associées à des données laser et RFID acquises

de manière synchrone. Chacune d'entre elles met en jeu entre trois et cinq personnes au maximum. Ces données synchrones ont été acquises depuis nos plateformes dans diverses conditions : (i) robot statique dans l'environnement (séquence $\#1$, figure 4.3(a)), (ii) robot dynamique avec une trajectoire prédéfinie (séquence $\#2$, figure 4.3(b)) pour évaluer les performances de notre démarche indépendamment de toute tâche robotique. A notre connaissance, il n'existe aucune base publique exploitant des données issues de différents capteurs. De plus, la plupart des bases connues mettent en œuvre des applications de vidéo surveillance, reposant uniquement sur des jeux d'images.

Rappelons que nos évalutations préliminaires visent à valider notre stratégie de fusion de données dans le cadre d'un suivi multi-personnes, *i.e.* au sein d'un filtre particulaire $\mathrm{RJ-MCMC}$. Notre base de données, dont certaines images clés sont présentées en figure 4.3, montre la diversité des situations mises en jeu. Elles comprennent plusieurs personnes présentes simultanément autour du robot, des occultations ainsi que des entrées et des sorties de la zone d'observation du robot. Afin de vérifier la répétabilité du filtre, l'algorithme implémenté à été exécuté plusieurs fois sur chaque séquence pour l'ensemble des évaluations.

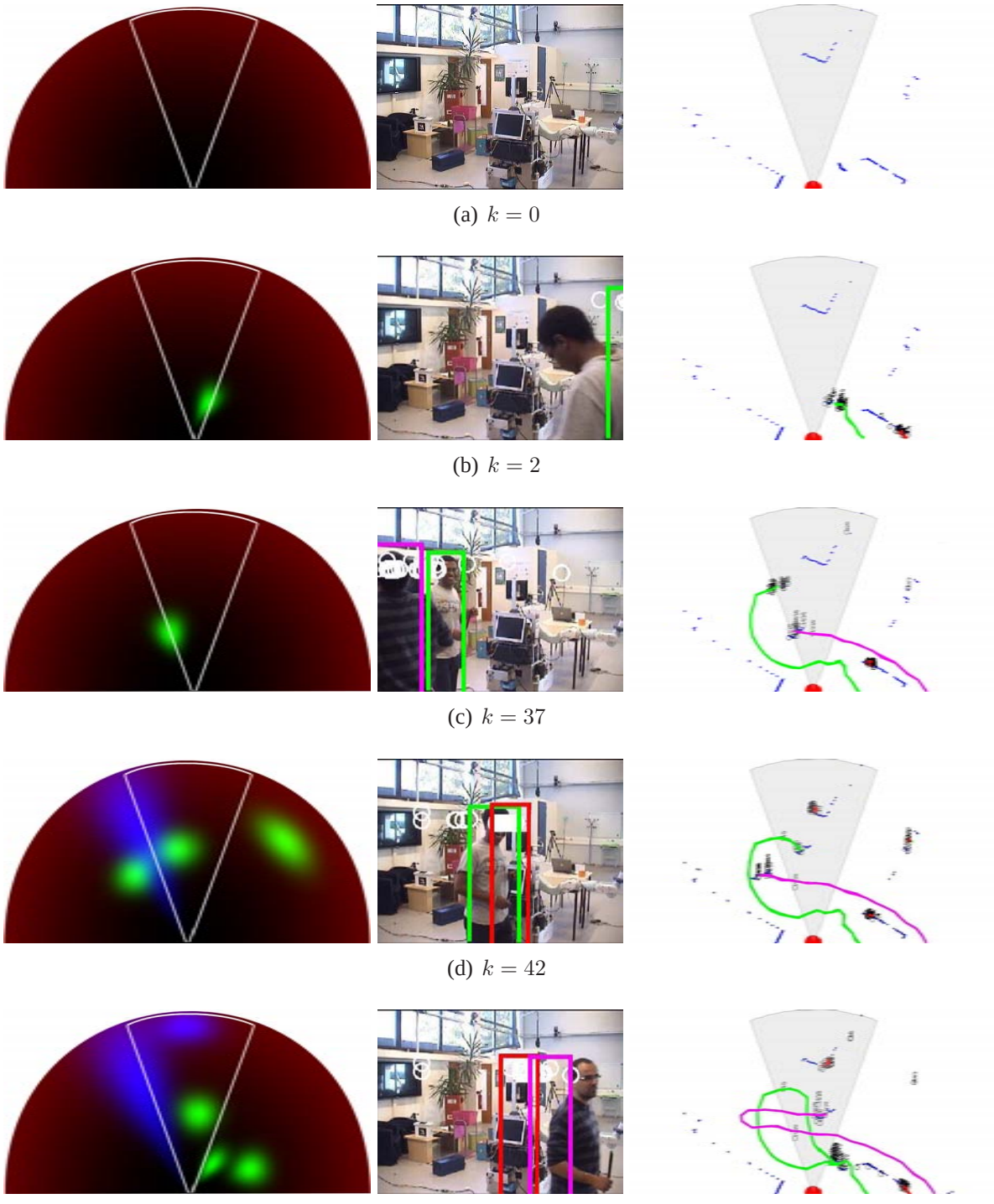
Tout d'abord, notre stratégie est illustrée sur la séquence $\#1$ *i.e.* robot arrêté, plusieurs personnes passant devant le robot. La figure 4.4 détaille le comportement du filtre dans différentes situations. Les trajectoires des différentes cibles sont représentées. Les trajectoires en vert et magenta correspondent à des cibles réelles, alors que celles en rouge ne correspondent à aucune personne dans l'environnement *i.e.* des *fantômes*.

Au départ de la séquence (figure 4.4(a)), la scène est vide, *i.e.* aucune cible n'est présente. Après l'entrée d'une cible (figure 4.4(b)), on remarque que l'utilisation de la fonction de proposition fusionnant les données laser et vision ainsi que la connaissance *a priori* permet d'échantillonner un grand nombre de cibles sur les détections. Après plusieurs images ($k = 37$), deux cibles se trouvent dans la scène (figure 4.4(c)). On peut alors remarquer que, malgré le fait qu'une seule personne soit détectée par le laser, les deux cibles (modulo les quelques artefacts de l'environnement) sont localisées et identifiées avec succès. Au contraire, lorsque de fausses détections apparaissent (figure 4.4(d)), la fonction de mesure, décrite par l'équation 4.17, permet d'invalider un saut d'entrée incohérent. Ici, une fausse détection RFID, apparemment dû à des réflexions dans l'environnement est observable en bleu sur la figure 4.4(d). De même, la sortie effective d'une cible de la scène (figure 4.4(e)) n'intervient que quelques instants aprés sa sortie réelle.

L'évaluation des performances quantitatives de la stratégie implémentée a été réalisée sur la base de données présentée figure 4.3. Deux critères classiquement utilisés dans la littérature [Bardet and Chateau, 2008] pour évaluer ce type d'algorithme ont été choisis : le taux de suivi, et le taux de fantômes.

– Le taux de suivi $\mathcal{R}_{suivi,k}$ permet d'évaluer la capacité de l'algorithme à effectivement suivre une cible, *i.e.* les vrais-positifs. Il est défini à l'instant k par :

$$\mathcal{R}_{suivi,k} = \frac{1}{\mathbf{I}_k} \sum_j \delta_k^s(j) \tag{4.18}$$



(a) $k = 0$

(b) $k = 2$

(c) $k = 37$

(d) $k = 42$

(e) $k = 75$

FIG. 4.4 – Images clés de la séquence #1. Première colonne : Carte de saillance des détections RFID (bleu) et laser (vert) et de la connaissance *a priori* (rouge). Deuxième colonne : Images associées issues de la caméra et projection du nuage de particules échantillonnées dans le plan image. Troisième colonne : représentation du nuage de particules échantillonnées et des trajectoires des cibles dans le plan du sol. Le champ de vue de la caméra est représenté par la zone grisée.

où $\delta_k^s(j) = 1$ si la personne $\mathbf{X}_{k,j}$ est correctement suivie à l'instant k, 0 sinon, et $\mathbf{I}_k$ est le nombre réel de personnes présentes dans la scène à l'instant k.

– Le taux de fantômes $\mathcal{R}_{ghost,k}$ représente les échecs du suivi relatifs aux fausses détections, *i.e.* les faux positifs. Il est défini à l'instant k comme suit :

$$\mathcal{R}_{ghost,k} = \frac{1}{\mathbf{I}_k} \sum_j \delta_k^g(j) \tag{4.19}$$

où $\delta_k^g(j) = 1$ si la cible ne correspond à aucune personne réelle et est présente dans plus de 3 cycles successifs.

De plus, afin de vérifier la pertinence des lois de proposition décrites dans la section 4.4.3, nous proposons de calculer le ratio $\mathcal{R}_{eff,k}$ relatif au nombre d'évènements effectivement acceptés lors de l'étape 8 de l'algorithme 4.2. Ce ratio est décrit par :

$$\mathcal{R}_{eff,k} = \frac{1}{N_B + MN_k} \sum_{i=1}^{N_B+MN} acc(\mathbf{X}_k^{(n)',i}), \tag{4.20}$$

où $acc(\mathbf{X}_k^{(n)',i}) = 1$ si la proposition $\mathbf{X}_k^{(n)',i}$ à l'itération i de la chaîne est acceptée à l'étape 8 de l'algorithme 4.2.

La table 4.2 présente des évaluations quantitatives sur notre base de données synchrones. Les résultats présentés donnent les taux moyens et leur écart-type pour des évaluations utilisant : (1) uniquement les données laser (table 4.2(a)), (2) les données laser et vision (table 4.2(b)), (3) les données laser, vision et RFID (table 4.2(c)). Chaque séquence a été évaluée séparément afin de comparer les différences de comportement du filtre dans des contextes impliquant un robot statique ou non.

Les premiers résultats issus de l'implémentation de l'algorithme $\mathrm{RJ-MCMC}$ pour le suivi multi-sensoriel multi-personnes sont encourageants. En effet, le taux moyen de suivi est $\mathcal{R}_{suivi} = 0.93$ pour la méthode de fusion de données laser/vision/RFID, ce qui signifie qu'une cible réelle est correctement suivie la plupart du temps alors que le système n'utilisant que les données laser a un taux moyen de suivi de $\mathcal{R}_{suivi} = 0.75$, du fait que le laser ne permet pas de détecter systématiquement une cible *i.e.* lorsqu'une seule jambe est détectée. Le comportement du filtre reste cependant assez stable d'une évaluation à l'autre ($\sigma_{\mathcal{R}_{suivi}} = 0.09$).

Les limitations portent sur les faux positifs dont le taux reste non négligeable ($\mathcal{R}_{ghost} = 0.15$) pour le système n'utilisant que les données laser. En effet, comme expliqué précédemment, les fausses détections entraînent l'apparition de cibles fictives (ou *fantômes*) dans la scène. L'apport de mesures complémentaires (au niveau des détections) comme la vision ou le RFID permet de réduire ce taux ($\mathcal{R}_{ghost} = 0.06$). De plus, un réglage plus fin des différents paramètres libres pourrait améliorer sensiblement ces résultats.

Une observation plus fine des résultats, notamment ceux concernant la fusion de données (laser/vision table 4.2(b) *vs.* laser/vision/RFID table 4.2(c)), permet de mettre en évidence l'apport de la fusion multi-capteurs *i.e.* le capteur visuel et le capteur RF. En effet, il semble que, dans

(a) Laser

Séquence testée	$\mathcal{R}_{suivi}$		$\mathcal{R}_{ghost}$		$\mathcal{R}_{eff}$
	μ	σ	μ	σ	
Séquence $\#1$	0.74	0.07	0.15	0.04	0.81
Séquence $\#2$	0.76	0.12	0.16	0.03	0.77
Total	**0.75**	**0.10**	**0.15**	**0.04**	0.79

(b) Laser + vision

Séquence testée	$\mathcal{R}_{suivi}$		$\mathcal{R}_{ghost}$		$\mathcal{R}_{eff}$
	μ	σ	μ	σ	
Séquence $\#1$	0.86	0.08	0.10	0.03	0.87
Séquence $\#2$	0.89	0.11	0.11	0.02	0.85
Total	**0.87**	**0.09**	**0.10**	**0.02**	0.86

(c) Laser + vision + RFID

Séquence testée	$\mathcal{R}_{suivi}$		$\mathcal{R}_{ghost}$		$\mathcal{R}_{eff}$
	μ	σ	μ	σ	
Séquence $\#1$	0.92	0.09	0.05	0.03	0.89
Séquence $\#2$	0.95	0.10	0.07	0.02	0.88
Total	**0.93**	**0.09**	**0.06**	**0.02**	0.89

TAB. 4.2 – Résultat des évaluations quantitatives du suivi multi-cibles.

notre contexte, l'ajout des détections visuelles et RFID permettent d'améliorer significativement les performances au regard de la méthode n'utilisant que les données laser. De plus, la multiplication des détecteurs permet de d'accroître significativement le nombre de cibles acceptées lors de la génération de la chaîne ($\mathcal{R}_{eff} = 0.89$ pour la stratégie (c) contre 0.86 pour la stratégie (b) et 0.79 pour la stratégie (a)). Il est donc possible de réduire les étapes de *burn in* et *thin out* en conséquence, car la chaîne de Markov converge beaucoup plus rapidement vers la distribution *a posteriori* grâce à l'utilisation de fonctions de proposition multi-sensorielles.

4.6 Conclusion

Ce chapitre a présenté nos développement réalisés en fin de thèse sur le suivi multi-cible depuis une plateforme mobile. Ces développements sont un premier pas vers l'évitement de cibles mobiles et plus généralement dans notre contexte de suivi de personnes en environnement encombré tel que nous l'avons défini au chapitre 1.

Les évaluations présentées ici étant préliminaires, la principale contribution de ce chapitre consiste en l'utilisation de données multi-sensorielles hétérogènes au sein d'un filtre à particules multi-cibles basé sur une méthode $\mathrm{RJ-MCMC}$. Les premières évaluations semblent prometteuses quant à l'utilisation d'une fonction de proposition multimodale dans la gestion des différents sauts. En effet, la fusion des données issues des différents capteurs permet de focaliser

l'échantillonnage des différentes cibles sur les zones pertinentes de l'espace d'état. L'observation des résultats quantitatifs permet de supposer une convergence plus rapide de la chaîne de Markov vers la distribution de probabilité à échantillonner *i.e.* une diminution du nombre d'itérations en phases de *burn in* et *thin out*. Des évaluations permettant de confirmer cette hypothèse et de quantifier les gains en temps de calcul sont en cours.

A l'instar du filtre particulaire implémenté au chapitre 3, le filtre particulaire $\mathrm{RJ-MCMC}$ reste robuste aux occultations sporadiques et aux artefacts de l'environnement, même si le nombre de "fantômes" peut être amélioré.

Bien évidemment, plusieurs travaux et investigations sont en cours concernant le suivi multi-sensoriel multi-personnes depuis une plateforme mobile.

Tout d'abord, il parait intéressant d'étudier une fonction de mesure combinant différents attribus. En effet, le champ de notre caméra étant très restreint, il ne nous est pas possible d'utiliser les mesures images persistantes définies au chapitre précédent *i.e.* une distribution de couleur. Or elle nous permettrait de différencier les différentes cibles de manière plus robuste lors des occultations, ce que ne permet pas le laser. En conséquence du point précédent, il serait judicieux de changer la focale de notre capteur visuel afin de couvrir un champ de vue plus large. Dans cette optique, pourquoi ne pas utiliser une caméra panoramique ou de type "fish-eye" pouvant couvrir un angle de vue jusqu'à $180°$.

De même l'apport de mesures plus robustes basées sur la vision permettrait de s'affranchir du RFID, car il est difficilement concevable d'équiper l'ensemble des passants d'un badge RFID. Dans notre contexte, seul l'utilisateur serait alors équipé d'un badge RF.

Chapitre 5

Intégrations et évaluations robotiques

La finalité de nos travaux est de permettre à un robot assistant d'interagir de manière naturelle avec son utilisateur par son positionnement aux alentours de ce dernier et de partager harmonieusement l'espace avec les passants. En effet, si le robot conserve en permanence une distance sociale avec son utilisateur, il sera en mesure d'interagir plus facilement avec ce dernier lorsque le besoin s'en fera sentir.

Ce chapitre traite de l'intégration des différentes fonctionnalités décrites dans les chapitres précédents au sein de plateformes robotiques. Le but de ce chapitre est donc de démontrer l'utilité de telles fonctions dans le cadre d'interactions Homme / Robot par le biais de démonstrations complètes, exploitant aussi bien les capacités intrinsèques du robot que les fonctionnalités de perception de l'homme pour l'interaction Homme / Robot que nous avons présentées. En effet, bien qu'un grand nombre de travaux existent sur la perception de l'homme en milieu encombré (cf. sections 2.1.2, 2.2.2, 3.1, 4.1), peu d'entre eux ont abouti à des intégrations et évaluations dans un contexte environnemental dynamique et encombré. Pour notre part, et bien que celà représente un investissement conséquent en terme de temps de travail, nous attachons une grande importance à cette phase d'intégration qui permet de valider nos fonctionnalités et valoriser nos travaux. Nous visons à rendre nos modules aussi génériques que possibles pour leur permettre d'être utilisés sur différentes plateformes et cherchons démontrer la pertinence de nos fonctions au travers de démonstrations robotiques.

Ce chapitre est structuré comme suit. La section 5.1 rappelle les différents contextes d'application de notre système au travers des plateformes et des outils utilisés pour son intégration. La section 5.3.1 détaille l'intégration du système de suivi d'utilisateur au sein d'une tâche robotique de suivi de personne alors que la section 5.4 met en jeu le suivi multi-cible dans une tâche d'évitement de personnes. La section 5.5 conclut ce chapitre en rappelant nos contributions et en énonçant quelques perspectives.

5.1 Plateformes robotiques et contextes applicatifs

Etant donné notre cadre applicatif, la validation de nos travaux passe logiquement par leur intégration sur des plateformes robotiques et leur évaluation à travers différents scénarii. Nous

allons donc décrire ici les robots ainsi que les outils utilisés pour l'intégration.

5.1.1 Plateformes robotiques

Cette sous-section a pour but de présenter les plateformes robotiques sur lesquelles nous avons implémpenté et évalué les travaux décrits dans les chapitres précédents. Précisons que l'intégration de nos travaux sur plusieurs plateformes se justifie par notre volonté de généricité qui seule prouve que notre approche et nos modules sont robustes à un grand nombre de situations

Le robot-assistant LAAS : Rackham

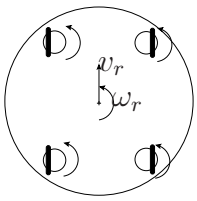
Rackham est un robot guide développé dans le but d'assister des personnes dans un musée. Tous ses capteurs permettent à Rackham d'agir en tant que robot de services dans un lieu public. En effet, il embarque de nombreux capteurs lui conférant de bonnes capacités d'interaction Homme / Robot. De plus, sa physionomie lui permet de naviguer aisément dans une foule de personnes alors que sa base non holonome est un atout pour l'interaction Homme / Robot du fait que ses capteurs font toujours face à la personne avec laquelle il interagit. Rackham a principalement été utilisé comme plateforme de test pour nos développements.

Rackham est une plateforme mobile B21r de iRobot (figure 5.1(a)). Il est constitué d'une base mobile de 52cm de hauteur et d'un corps cylindrique de 118cm intégrant deux PC et surmonté d'un mât. Ses équipements standards ont été étendus avec :

- un écran tactile,
- une paire de haut-parleurs,
- un gyroscope,
- une camera firewire montée sur un platine orientable dédiée à la perception de l'utilisateur,
- une caméra Pan-Tilt-Zoom analogique dédiée à la détection des passants,
- un laser avec un champ de vue de $180°$ orienté vers l'avant,
- le système RFID décrit en section 2.2.

Les déplacements de Rackham sont assurés par quatre roues orientables (figure 5.1(b)). L'orientation de ces roues, rigidement liées les unes aux autres, permet d'orienter le corps de Rackham alors que leur actionnement conjoint lui permet de se mouvoir.

(a) Rackham



(b) Actionneurs de Rackham

FIG. 5.1 – Détail de la plateforme Rackham.

Quatre valeurs peuvent être contrôlées pour permettre au robot de réaliser une tâche définie : les vitesses pan et tilt de la platine (ω_p, ω_t) permettant d'orienter la caméra, et les vitesses linéaire et angulaire de la plateforme (v_r, ω_r) permettant de déplacer le robot.

Dans nos évaluations, Rackham est alors utilisé pour réaliser des tâches robotiques relatives à son domaine d'application *i.e.* (i) accompagner l'utilisateur dans un lieu public, (ii) éviter les personnes se trouvant dans son voisinage immédiat.

Le caddie CommRob : Inbot

Inbot est un caddie dit "intelligent", développé dans le cadre du projet européen CommRob (cf chapitre 1. Cette plateforme (figure 5.2) est développée par FZI (Karlsruhe, Allemagne) et est utilisée dans diverses démonstrations avec des environnements de types supermarché. Ce cadre a été choisi pour les situations qu'il génère qui sont à la fois courantes et habituelles pour l'homme, mais qui se déroulent dans des environnements encombrés et fortement dynamiques. Le rôle du robot est alors de pouvoir remplir les mêmes missions qu'un caddie classique, à la différence près qu'il est motorisé. Les mouvements du robots peuvent alors êtres induits par une interaction physique au travers d'une poignée haptique ou bien par une interaction sans contact au travers de l'interface perceptuelle embarquée que nous proposons. Inbot a été majoritairement utilisé pour les besoins du projet CommRob, en tant que plateforme d'intégration.

Le robot est pourvu de deux PC industriels dont l'un est dédié aux mouvements du robot et l'autre à la perception de l'homme ainsi que d'un PC tactile dédié à l'interface Homme / Machine. De plus, pour l'interaction Homme / Robot, le robot est équipé entre autre de :

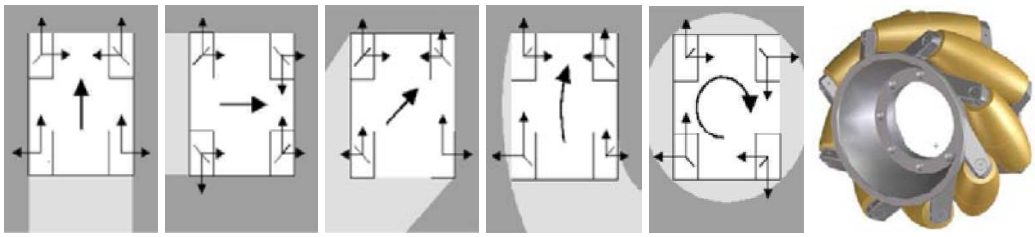
- une poignée haptique,
- deux lasers ayant un champ de vue de $270°$ chacun,
- le système RFID décrit en section 2.2,
- un système de caméras stéréo monté sur une platine orientable.

La plateforme Inbot est une base holonome capable de se déplacer dans toutes les directions. Ces mouvements sont obtenus par l'action conjointe des quatres roues indépendantes présentées figure 5.3(f). En fonction du sens et de la vitesse de rotation relative des roues, il est possible de généré différents mouvements dont les plus élémentaires sont présentés sur les figures 5.3(a)-(e).

FIG. 5.2 – Inbot.

Les déplacements de Inbot sont alors contrôlés au moyen de trois valeurs : les vitesses linéaires (v_{r_x}, v_{r_y}) permettant une translation du robot suivant les axes x (vers l'avant) et y (sur le côté), et la vitesse angulaire ω_r. De plus, à l'instar de Rackham, les mouvements de la platine orientable sont contrôlés grâce aux vitesses pan et tilt (ω_p, ω_t).

Dans nos évaluations, Inbot est alors utilisé pour réaliser les tâches robotiques relatives à son domaine d'application décrit en section 1.2 *i.e.* (i) le *following mode* (suivi de personne), (ii) le *guiding mode* (guidage de personnes).



(a) Avant (b) Côté (c) Diagonal (d) Courbe (e) Rotation (f) Roues de Inbot

FIG. 5.3 – Modes de déplacements de la plateforme Inbot.

5.1.2 Environnements de développements logiciels

La figure 5.4 rappelle l'architecture générale du système de perception proposée au chapitre 1. Chaque robot implémente alors ces différentes fonctionnalités au sein de l'architecture logicielle et matérielle qui le caractérise.

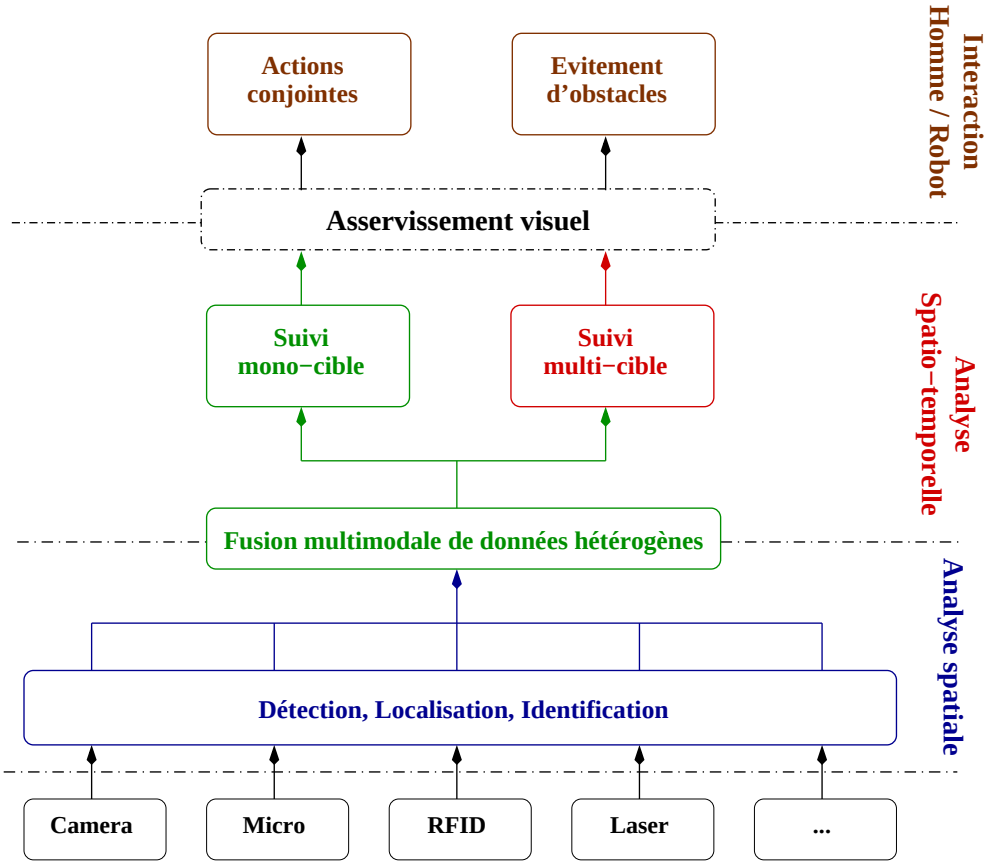


FIG. 5.4 – Description générique de l'architecture des fonctionnalités perceptuelles à intégrer.

GenoM

Dans le domaine de la conception d'architecture de contrôle pour l'autonomie des robots, le pôle Robotique et Intelligence Artificielle du LAAS-CNRS utilise et développe l'outil $G^{en}{}_oM$ (pour Générateur de modules ou *GENenrator Of Modules*) [Alami et al., 1998]. Ce dernier permet de générer automatiquement des composants logiciels temps-réel en s'appuyant sur un langage de description de composants (interfaces, propriétés temporelles, incompatibilités entre services, etc...) et un langage de programmation (C/C++). Grâce à cette description, la partie algorithmique du composant est, entre autres, interfacée automatiquement aux couches de communication.

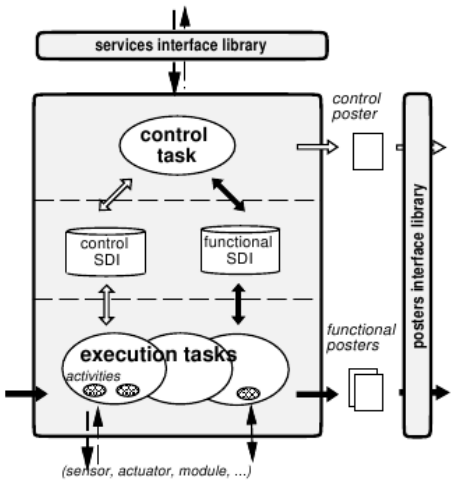
Le superviseur [Clodic et al., 2005] est une couche de haut niveau destinée à se servir des informations renvoyées par les différent modules fonctionnant en parallèles sur le robot afin de prendre des décisions concernant le comportement de ce dernier. Ses prises de décisions se traduisent par l'envoi de requêtes à exécuter aux différents modules concernés par la tâche en cours.

Chaque module (figure 5.5(a)) décrit :

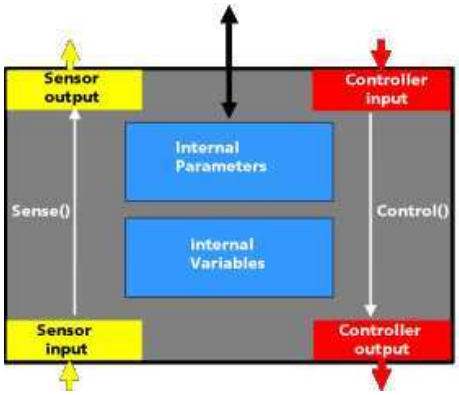
1. une (ou plusieurs) *tâche d'exécution* permettant l'exécution d'une routine principale,
2. des *requêtes* permettant d'implémenter une interface dédiée à la réception de commandes depuis un superviseur, un autre module ou un utilisateur,
3. des *posters* exportant les données issues du module en question et nécessaires (i) à l'exécution d'autres modules, (ii) à la compréhension du contexte général par le superviseur.

MCA2

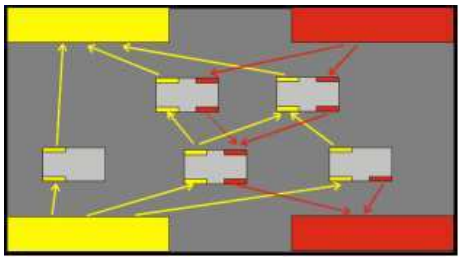
MCA2 (ou *Modular controller architecture version 2* pour Architecture de contrôle modulaire) est un logiciel ayant pour but de produire des programmes de contrôle pour les robots autonomes. MCA est un système modulaire, temps-réel pour le contrôle de robots ou autre type de matériels. L'utilisation de compo-



(a) Description générique d'un module $G^{en}{}_oM$



(b) Description générique d'un module MCA



(c) Description générique d'un groupe MCA

FIG. 5.5 – Détail schématique des différentes environnements de développement.

sants logiciels portables est un aspect important pour le développement et la réalisation de robots. Des composants réutilisables avec une interface standardisée permettent une extension plus aisée des architectures. MCA2 permet donc de réaliser de tels composants logiciels.

Chaque module est structuré de la même manière (figure 5.5(b)). Il comprend cinq interfaces : quatre d'entre elles permettent la connection vers d'autres modules situés "au-dessus" ou "au-dessous" dans l'architecture. Dans chaque direction, il existe une interface pour la réception de données (*input*) et une autre pour l'envoi de données (*output*). Une autre interface consiste à permettre la lecture et la modification de paramètres internes au module, même pendant son exécution. De plus, chaque module implémente deux fonctions : *sense()* et *control()*. Ces fonctions sont responsables du comportement du modules en réponse aux données reçues.

Au sein de MCA2, chaque module fourni donc une interface simple et normalisée. Les modules développés sont alors connectés par des "*edges*" permettant les échanges de données. De plus, plusieurs modules peuvent être combinés au sein d'un groupe qui agit comme un simple module mais avec un comportement plus évolué (figure 5.5(c)).

Ce type d'environnement de développement a été utilisé dans le cadre du projet CommRob.

5.2 Implémentation

5.2.1 La bibliothèque *libVision*

Dans le but de capitaliser l'ensemble des développements relatifs à la perception de l'homme depuis une plateforme mobile, nous avons regroupé les fonctionnalités perceptuelles décrites dans ce manuscrit au sein d'une bibliothèque logicielle. Cette bibliothèque se compose de différentes parties :

- Les primitives de détection : détecteur de visages de Viola *et al.*, détecteur visuel de personnes décrit en section 2.3.2, segmentation de pixels de couleur peau, détection laser de personnes (section 2.3.1), détection RFID (section 2.2), et d'autres détecteurs qui n'ont pas été utilisés dans nos travaux, tels que la segmentation de zones en mouvements ;
- Les fonctions dédiées au filtrage : algorithmes SIR, CONDENSATION, ICONDENSATION, MCMC et RJ − MCMC ainsi que les fonctions telles que le rééchantillonnage des particules et le calcul des estimés MMSE et MAP ;
- Les fonctions de mesures : les mesures de distributions de couleurs et de contour décrites en section 3.5, ainsi que la mesure laser décrite en section 4.4.4 ;
- Les algorithmes de classification et d'apprentissage : apprentissage de bases ACP et AFD, apprentissage SVM, fonctions de projection dans un sous-espace ainsi que les différents algorithmes de classifications décrits en section 2.1 ;
- Les algorithmes de suivi : constructions des cartes de saillances basées sur les détections, algorithmes d'échantillonnage et fonctions de mesures globales ;
- Les fonctions géométriques : triangulation, projection de points 3D dans le plan image ... Ces fonctions, bien qu'elles n'aient pas été utilisées dans nos approches, ont été implémentées pour de possibles extensions au 3D.

Cette bibliothèque de fonctions perceptuelles a été conçue dans le but d'être étendue au fur et à mesure des développements. Elle contient aussi une section de programmes de tests. L'utilité d'une telle bibliothèque réside dans le fait qu'elle permet de tester les fonctions et algorithmes sur des séquences hors-ligne avant d'intégrer ces mêmes fonctions au sein d'une architecture plus spécifique comme il nous a fallut le faire pour $G^{en}{}_{o}M$ et MCA2.

5.2.2 Intégration sur le robot Rackham

Dans les évaluations qui vont suivre, nous nous focalisons sur les modules présentés dans la figure 5.6, *i.e.* `HumTrack`, `HumPos`, `RFID` et `Visuserv` qui implémentent respectivement l'identification et le suivi de l'utilisateur, la détection et le suivi multi-personnes, la localisation de badges RFID et la boucle d'asservissement visuel. La fusion des données issues des modules `Camera` et `RFID` est faite au sein du module `HumTrack` du fait que la vision constitue le capteur "central" du processus alors que la fusion de données nécessaire au suivi multi-cibles (*i.e.* caméra et laser) est faite au niveau du module `HumPos`. Les déplacements de la caméra ainsi que du robot sont calculés par `Visuserv` qui contrôle les deux modules effecteurs *i.e.* `Platine` et `Rflex`. Ces modules ont été implémentés dans $G^{en}{}_{o}M$ à l'aide de la bibliothèque OpenCV utilisée pour les primitives bas niveau *e.g.* extraction de contours, détection de visages et interfacés à notre bibliothèque libVision. L'ensemble du système fonctionne à une fréquence moyenne de 6 Hz.

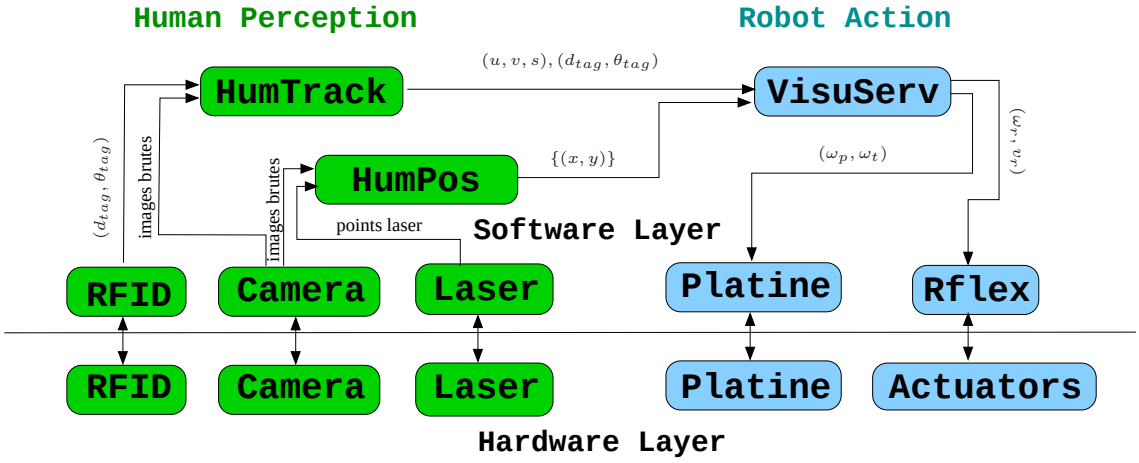


FIG. 5.6 – Modules impliqués dans l'intégration des fonctionnalités décrites.

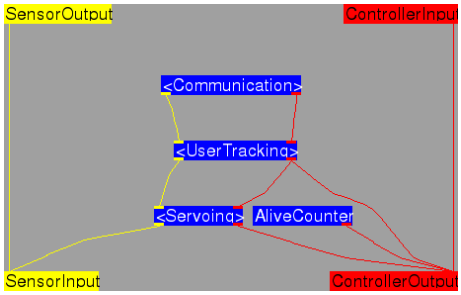
Cette architecture sera mise en œuvre lors des évaluations relatives au tâches robotiques de suivi et d'évitement de personnes.

5.2.3 Intégration sur le trolley Inbot

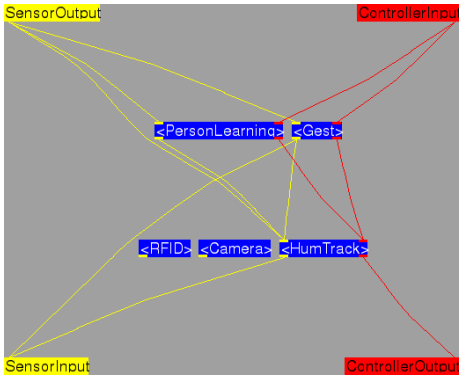
Dans le cadre du projet CommRob, les modules nécessaires à l'interaction Homme / Robot ont été développésau moyen de MCA2. Les modules mis en œuvre ont été regroupés en

différent niveau suivant leur fonctionnalités (figure 5.7(a)). Ainsi, coexistent (1) une couche de `Communication` visant à recevoir des commandes et transmettre des évènements via une connexion TCP, (2) une couche `User Tracking` (figure 5.7(b)) implémentant les algorithmes de perception sur l'homme, et (3) une couche `Servoing` (figure 5.7(c)) permettant de commander la platine et le robot en fonction des données issues de la couche de suivi.

A l'instar de l'architecture mise en place sur le robot Rackham, le module `HumTrack` fusionne les données issues des modules `Camera` et `RFID`. Pour les besoins du projet, d'autres modules ont été développés mais ne seront pas détaillés dans nos évaluations tels que `Gest`, dédié au suivi et à l'interpétation de gestes, et `PersonLearning` réalisant l'apprentissage en ligne de l'utilisateur d'après les techniques d'apprentissage de visages décrites au chapitre 2. Le groupe `UserTracking` est commandé contrôlé par deux commandes (*ControllerInput*) permettant de mettre en marche le suivi de personnes (`AttachUser`) ou de l'arrêter (`DettachUser`). Ce même groupe exporte différentes données (*ControllerOutput*) telles que la position de l'utilisateur dans l'image, la position approximative du badge RFID. De plus, le groupe `UserTracking` utilise les données de position de la platine (*SensorInput*) et exporte l'état du suivi donnant une information sur le contact visuel avec l'utilisateur (*SensorOutput*).

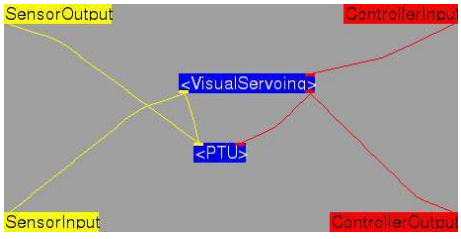


(a) Architecture générale des modules MCA2.



(b) Détail des modules du groupe `User Tracking`.

Le groupe de modules `Servoing` regroupe lui, les modules relatifs à l'asservissement visuel et à la commande des actionneurs du robot, à savoir `VisuServ` et `PTU`. Le module `PTU` contrôle les mouvements de la platine alors que le module `VisuServ` calcule les différentes vitesses des actionneurs en focntion des données qu'il reçoit (*ControllerInput*) du groupe `UserTracking`. Les commandes des actionneurs du robot sont exportées vers d'autres modules de l'architecture (*ControllerOuput*).



(c) Détail des modules du groupe `Servoing`.

L'ensemble des modules présentés ci-dessus ont été développés pour les besoins du projet et sont basés sur les bibliothèques OpenCV et libVision.

5.3 Asservissement visuel pour le suivi de l'utilisateur

Nous allons dans un premier temps traiter du problème de suivi de personne. Dans ce but, nous utilisons les données issues du traqueur défini au chapitre 3 ainsi que celles issues du système RFID défini en section 2.2. Nous allons d'abord présenter la stratégie de contrôle utilisée avant de détailler les différentes lois de commande choisies. Des évaluations, aussi bien qualitatives que quantitatives, seront ensuite menées afin de valider l'intégration de la fusion de données hétérogènes pour le suivi multimodal de l'utilisateur. Bien que le travail présenté dans la section 5.3.1 n'ait pas été réalisé dans le cadre de cette thèse (travaux de Noureddine Ouadah), sa présentation est nécessaire à la compréhension des évaluations de la tâche robotique globale qui vont suivre.

5.3.1 Loi de commande basée capteurs pour la tâche de suivi de personne

Notre but est de calculer les vitesses des actionneurs de chaque robot, *i.e* v_r, ω_r, ω_p, ω_t, afin que le robot puisse réaliser sa tâche de manière efficace et sûre. Plusieurs stratégies de contrôle sont disponibles dans la littérature. Dans notre cas où la vision est considérée comme le capteur principal, il paraît naturel de considérer les méthodes d'asservissement visuel [Espiau et al., 1992; Corke, 1996]. Ces techniques permettent de commander un grand nombre de systèmes robotiques grâce aux images issues d'une (ou plusieurs) caméra(s). Cependant, même si ces techniques sont utilisées dans un grand nombre d'applications, la littérature ne présente que peu de travaux concernant le problème d'un asservissement multi-capteur appliqué à la perception de l'homme. Notre idée est d'utiliser à la fois les données issues du traqueur visuel et les données RFID pour construire notre loi de commande. Nous avons choisi de définir chacune des lois de commande séparément afin de découpler au mieux les différents degrés de liberté relatifs au positionnement de la caméra. L'analyse de la structure du robot montre que ω_r et ω_p, relatifs respectivement à la vitesse angulaire du robot et à la vitesse en pan de la platine, ont le même effet sur les mouvements de la cible dans l'image. Même si cette propriété peut être utilisée pour d'autres objectifs, *e.g.* l'évitement d'obstacles (traité en section 5.4), nous avons dans un premier temps choisi de fixer ω_p à zéro afin de contrôler la position horizontale de la cible dans l'image grâce à un seul contrôleur. De plus l'utilisation de ω_r au lieu de ω_p permet d'orienter l'ensemble de la plateforme vers la cible (plutôt qu'uniquement la caméra) ce qui améliore les conditions d'interaction du point de vue de l'utilisateur.

Notre but est donc de définir trois contrôleurs $(v_r, \omega_r, \omega_t)$ afin d'orienter la caméra et de faire bouger le robot, de telle manière que la personne cible est toujours dans le champ de vue du robot à une distance sociale du robot. A cet effet, nous utilisons les données issues du traqueur *i.e.* les coordonnées de la personne cible dans l'image (u_{gc}, v_{gc}) et son échelle associée s_{gc} qui caractérise vaguement la distance Homme / Robot. De ces données, nous pouvons définir une fonction d'erreur E_{ptv} à faire tendre vers zéro :

$$E_{ptv} = \begin{pmatrix} E_u & E_v & E_s \end{pmatrix}^T = \begin{pmatrix} u_{gc} - u^* & v_{gc} - v^* & s_{gc} - s^* \end{pmatrix}^T,$$

où u^* et v^* représentent le centre de l'image et s^* une échelle prédéfinie correspondant à une distance sociale acceptable anotée d_{suivi}.

E_u représente l'erreur en abscisse dans l'image et peut alors être régulée à zéro en agissant sur la vitesse angulaire du robot ω_r, E_v correspond à l'erreur en ordonnée dans l'image est peut donc être régulée à zéro en agissant sur la vitesse en tilt de la platine ω_t et E_s correspond à l'erreur en échelle régulée à zéro par l'action sur la vitesse linéaire du robot v_r. Nous pouvons alors définir trois contrôleurs PID (pour Proportionnel, Intégral, Dérivé) comme suit :

$$\begin{cases} \omega_r &= K_{pp}E_u + K_{ip}\int E_u dt + K_{dp}\frac{dE_u}{dt} \\ \omega_t &= K_{pt}E_v + K_{it}\int E_v dt + K_{dt}\frac{dE_v}{dt} \\ v_r &= K_{pv}E_s + K_{iv}\int E_s dt + K_{dv}\frac{dE_s}{dt} \end{cases} \tag{5.1}$$

Les gains de contrôle (K_{pp}, K_{ip}, K_{dp}) (respectivement, (K_{pt}, K_{it}, K_{dt}) and (K_{pv}, K_{iv}, K_{dv})) sont déterminés empiriquement et assurent la stabilité du système.

Cependant, ces lois ne peuvent être utilisées uniquement lorsque le contact visuel entre l'utilisateur et le robot existe *i.e.* lorsque l'utilisateur est dans le champ de vue du robot. Lorsque la personne cible est perdue, il n'est alors plus possible d'appliquer ces lois de commande. Nous proposons alors d'utiliser les informations issues du système RFID, à savoir la distance d_{tag} et l'orientation θ_{tag} de l'utilisateur afin de contrôler le robot. L'idée est alors d'orienter la caméra en direction du badge afin que le traqueur puisse retrouver l'utilisateur dans le champ de vue de la caméra. Le comportement du robot est alors décrit dans la figure 5.7. Dans ce but, nous appliquons simplement une vitesse angulaire constante ω_r^0 en ω_r. La vitesse en tilt ω_t est commandée dans le but de ramener l'angle de la platine à sa position d'origine. De plus, nous appliquons une vitesse linéaire au robot dépendant de d_{tag} et θ_{tag}. Dans ce but, nous essayons de satisfaire la contrainte relative à la conservation de la distance sociale d_{suivi}[1], malgré la perte du contact visuel. Le robot est alors maintenu dans le proche voisinage de l'utilisateur afin de faciliter la reprise du contact visuel. Ensuite, lorsque le contact visuel avec la personne cible est à nouveau possible, la stratégie de contrôle basée sur les données du traqueur définie précédemment reprend. Il faut alors noter que la continuité de la loi de commande est assurée lors du changement de comportement entre le contrôle basé sur les RFID et celui basé sur la vision. En effet, la vitesse linéaire du robot est modifiée progressivement afin d'atteindre la valeur désirée. De même, la continuité des lois relatives à ω_r et ω_t n'est pas gérée de manière explicite car toutes deux correspondent à des systèmes suffisamment réactifs.

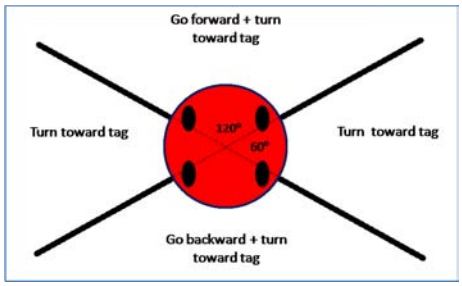


FIG. 5.7 – Comportement du robot basé sur les détections RFID.

5.3.2 Expérimentations sur Rackham et discussions associées

Une campagne d'expérimentation a été réalisée dans notre salle robotique (4×5 m^2) en présence de nombreuses personnes afin de valider l'approche proposée. Rappelons que notre

[1]La distance d_{tag} fournie par le système RFID n'étant pas très précise.

objectif est d'effectuer une tâche consistant à accompagner une personne novice, équipée d'un badge RFID, dans un lieu dynamique, naturel et encombré en respectant les contraintes imposées : (i) l'utilisateur doit toujours être centré dans l'image et (ii) une distance $d_{suivi} = 2$m doit être maintenue entre l'utilisateur et le robot. Pendant ces évaluations, pour des raisons de sécurité, les vitesses linéaires et angulaires du robot ont été limitées à respectivement 0.4m.s^{-1} et 0.6rad.s^{-1}. Ces valeurs sont compatibles avec les déplacements de l'utilisateur[2].

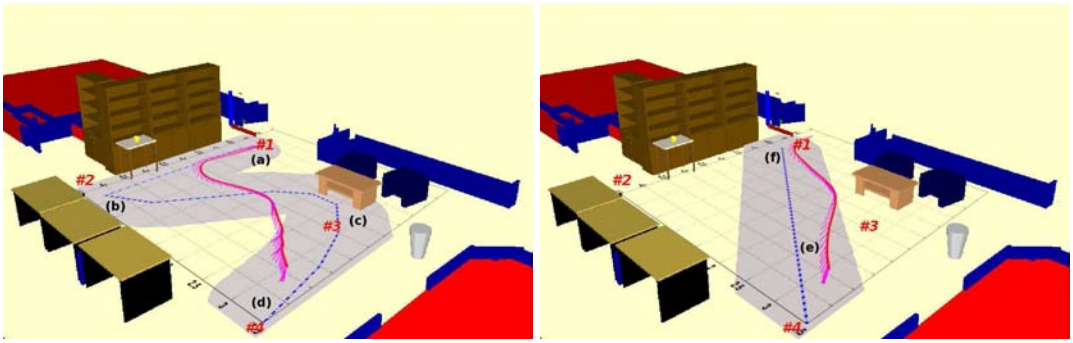


FIG. 5.8 – Exécution classique d'une tâche robotique de suivi sans obstacles : la ligne pleine rouge (resp. pointillée bleue) représente les positions du robot (resp. de l'utilisateur), alors que les lignes violettes montrent la direction de la caméra.

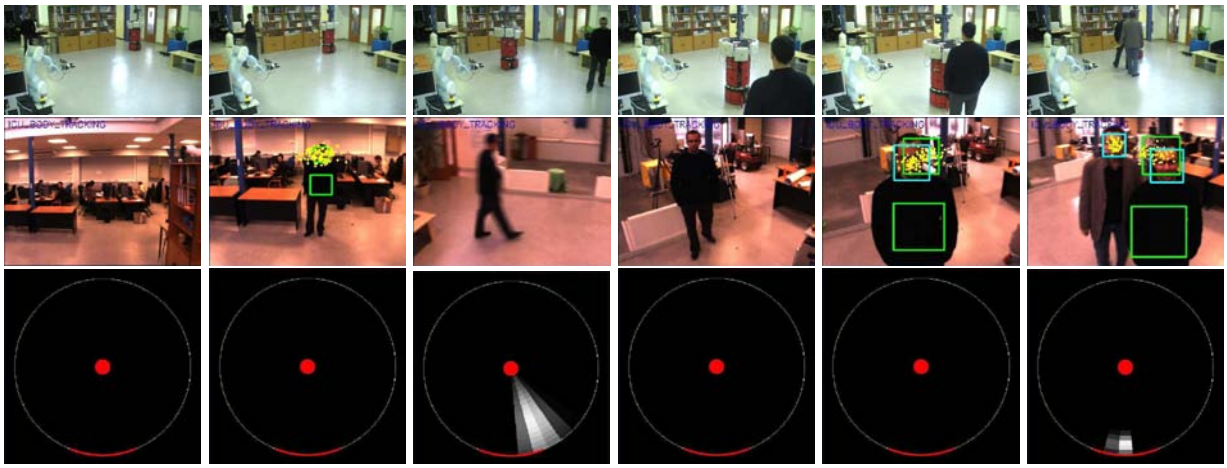
La figure 5.8 décrit l'environnement dans lequel évolue Rackham, la trajectoire prédéfinie de l'utilisateur ainsi que la trajectoire du robot résultant de l'exécution de la tâche robotique dans un tel environnement. De nombreuses séries de tests ont été effectuées selon le scénario suivant : un visiteur entre dans la salle au niveau du point $\#1$ et prend un badge RFID. L'utilisateur va ensuite vers le point $\#2$ avant de continuer vers le point $\#4$ *via* le point $\#3$. Il attend ensuite le robot avant de retourner au point $\#1$ afin de quitter la salle. Durant tout le trajet de $\#1$ à $\#4$, Rackham devra accompagner la personne en restant derrière elle, alors que durant le trajet retour (de $\#4$ à $\#1$) Rackham restera devant l'utilisateur en tentant de conserver au mieux une distance sociale acceptable *i.e* d_{suivi}.

Dans ces conditions nominales, la figure 5.9 détaille les images clés du flux vidéo ainsi que les sorties du traqueur et les cartes de saillance RFID associées. La figure 5.10 montre alors les différents signaux des modules `HumTrack`, `RFID` et `Visuserv` à savoir :

- les deux variables `HumTrack` et `RFID` qui prennent la valeur 1 lorsque la cible est détectée respectivement dans l'image et par le système RFID,
- l'angle θ_{tag} et la distance d_{tag} mesuré par le système RFID,
- les trois valeurs (v_r, ω_r, ω_t) calculées par le module `Visuserv` et envoyées au robot.

Après l'initialisation de la mission (a), le robot focalise son attention sur la personne cible grâce aux données vidéo (b). Les quatre étapes du scénario sont alors exécutées. Entre les points $\#1$ et $\#2$, le contact avec la cible est maintenu grâce au système de vision. Les lois de commande

[2]Si l'utilisateur avance plus vite que le robot et est perdu, le robot s'arrête par mesures de sécurité.



(a) (b) (c) (d) (e) (f)

FIG. 5.9 – Exemple d'une séquence. Notons que la position en azimut de la platine est représentée par un arc de cercle rouge sur la carte RFID. Les carrés bleus et verts représentent respectivement les détections de visages et l'estimée MMSE alors que les points jaunes représentent le nuage de particules avant rééchantillonnage.

sont alors calculées en fonction des données fournies par le traqueur. La cible est centrée dans l'image alors que le robot ajuste sa distance à l'utilisateur au moyen de l'échelle du modèle. Entre les points $\#2$ et $\#4$ (c), la cible sort du champ de vue de la caméra, ce qui induit un échec du traqueur visuel. Les lois de commande sont alors calculées grâce aux information RFID (d_{tag}, θ_{tag}) pour permettre au robot de faire face à la personne cible et converger vers cette dernière jusqu'à ce que d_{tag} soit proche de d_{suivi}. La tâche de suivi de personne est donc maintenue malgré la perte du contact visuel alors que le système RFID guide la caméra afin de récupérer le contact visuel (d). L'identification visuelle de l'utilisateur permet alors de réinitialiser le traqueur (e) pendant que l'utilisateur retourne vers le point $\#1$ (f). Pendant cette phase du scénario, la trajectoire du robot rencontre celle de l'utilisateur. Comme attendu, dans le but de préserver une distance sociale, le robot recule. Dans ces conditions nominales, la tâche d'accompagnement de l'utilsateur est réalisée avec succès. L'erreur de suivi moyenne durant toute la mission est de 0.08m malgré le faible apport du système RFID.

Nous avons, dans un deuxième temps, réalisé une série de tests pour effectuer des évaluations aussi bien qualitatives que quantitatives du scénario détaillé ci-dessus. Ces évaluations ont été effectuées avec l'aide de personnes dont la plupart ne connaissaient pas le système qui était testé. Le scénario a été effectué 10 fois pour chaque utilisateur en augmentant le nombre de personnes présentes autour du robot afin de générer des problèmes d'occultations durant la mission. Le but de ces évaluations étant de quantifier la robustesse de la tâche robotique d'accompagnement, et non d'évitement, les passants occultent la personne cible mais ne perturbent pas la trajectoire du robot. Cette deuxième tâche sera évaluée plus loin. La figure 5.11 décrit la configuration typique de l'environnement lors de telles évaluations. Plusieurs personnes passent entre Rackham et son

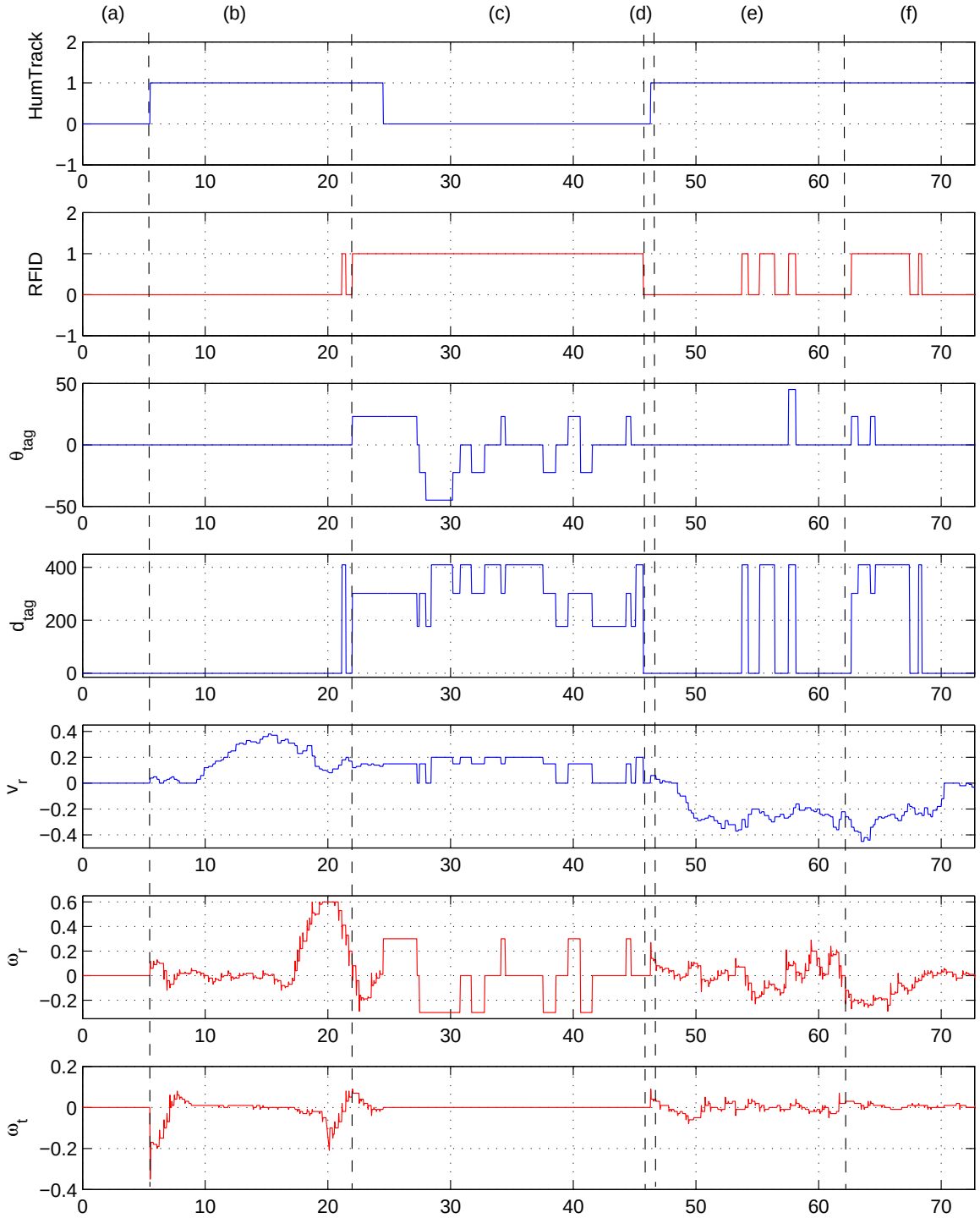


FIG. 5.10 – Synchronisation des flux de données entre les différents modules.

utilisateur, marchent à côté de ce dernier ou même devant Rackham pendant un long moment. La perte de la cible provoque alors l'arrêt du robot et une réinitialisation du traqueur.

Plusieurs séries de 10 réalisations ont été menées. Les performances du système sont évaluées sur l'ensemble des expérimentations et quantifiées par :

- le *Ratio de Contact Visuel* (RCV) défini par le rapport entre le temps où l'utilisateur est présent dans le champ de vision du robot et le temps total de la mission. Cet indicateur mesure indirectement la robustesse de l'algorithme de suivi aux artefacts tels que les occultations, les disparitions temporaires dues à la présence de passants.
- l'*Erreur de Guidage* définie par $E_{suivi} = |d - d_{suivi}|$ où d est la distance courante à l'utilisateur évalué à partir de l'échelle s de l'estimé dans l'image. Cette erreur mesure la capacité du robot à suivre une personne à une distance définie d_{suivi}.

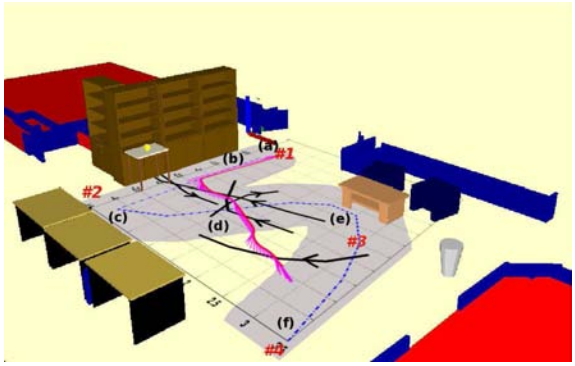


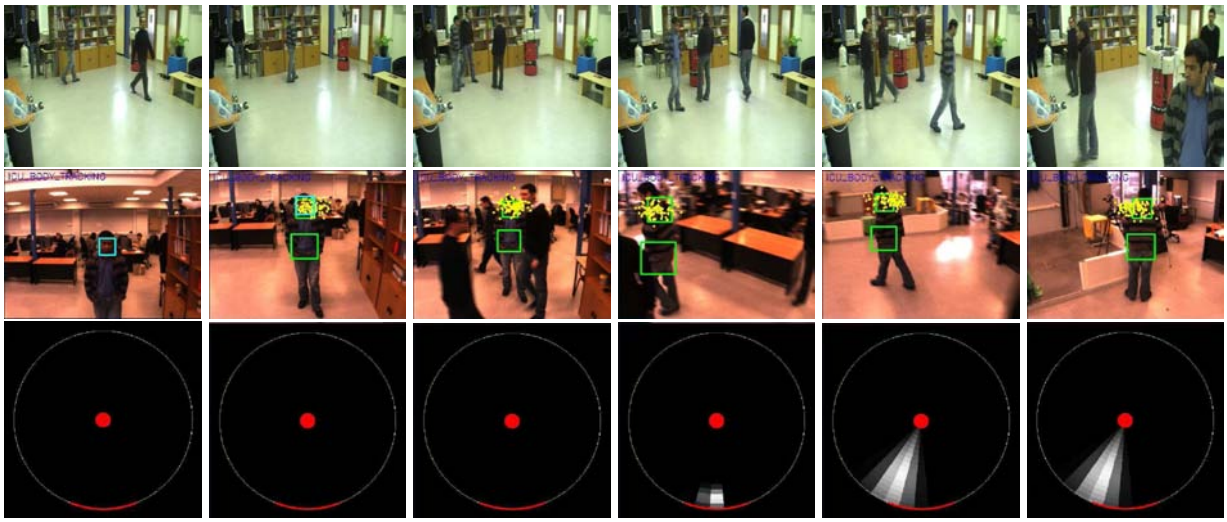
FIG. 5.11 – Exécution d'une tâche robotique de suivi avec obstacles : la ligne pleine rouge (resp. pointillée bleue) représente les positions du robot (resp. de l'utilisateur), alors que les lignes violettes montrent la direction de la caméra. Les flèches noires représentent les trajectoires des passants.

La figure 5.12 montre une séquence type des tests effectués[3] correspondant à l'environnement décrit dans la figure 5.11. Dans la plupart des cas, le système basé sur la vision seule rencontre des difficultés à se (ré-)initialiser après de longues occultations de la personne cible, alors que le système multimodal est capable de dépasser ce type de problèmes. Ainsi, le premier dispositif exécute 12% des missions avec succès, tandis que plus de 85% d'entre elles sont réussies par le système multimodal. Le tableau 5.1 détaille ces résultats avec un nombre croissant de passants au voisinage du robot.

Le Ratio de Contact Visuel reste constant pour les deux systèmes évalués, mais les résultats montrent l'efficacité du système multimodal. En effet, l'ajout du dispositif RFID permet de conserver la cible dans le champ de vision plus de $\mu = 85\%$ du temps de la mission malgré la présence de plus de 4 personnes autour du robot. La valeur importante de l'écart-type (noté σ) est en grande partie due aux déplacements aléatoires que nous avions demandés aux passants afin d'évaluer le système dans des conditions réalistes. Nous avons également mesuré l'erreur de guidage pendant ces tests. Sa valeur moyenne, autour de 0.10m, est un peu plus élevée que celle mesurée dans les conditions nominales puisqu'elle intègre les tests effectués avec des passants.

On ne peut alors que remarquer que l'ajout du RFID à la vision permet d'améliorer la robustesse du système aux artefacts de l'environnement tels que les occultations longues, les sorties du champ de vue de la caméra.

[3]La séquence complète est disponible à l'adresse suivante : `homepages.laas.fr/tgerma/RFIA`



(a) (b) (c) (d) (e) (f)

FIG. 5.12 – Séquence type pour l'évaluation du système. La première ligne montre la situation Homme / Robot courante, la deuxième montre le résultat du suivi et la troisième montre la carte de saillance du système RFID.

Système testé	Nombre de passants :								Total	
	1		2		3		4 et plus			
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Vision seule	0.21	0.11	0.22	0.02	0.18	0.05	0.22	0.06	0.21	0.04
Vision + RFID	**0.94**	**0.08**	**0.85**	**0.14**	**0.94**	**0.13**	**0.83**	**0.19**	**0.86**	**0.14**

TAB. 5.1 – Résultat du *Ratio de Contact Visuel* considérant de 1 à 4 passants.

5.3.3 Evaluations sur Inbot et discussions associées

L'aspect confidentiel du travail de certains partenaires ne nous permet pas de présenter en détail nos expérimentations. Il était notamment interdit de filmer les évaluations durant la présentation des scénarii. Cependant, voici une rapide description des tests effectués dans le cadre du projet.

Les scénarii étudiés correspondent à ceux décrits dans la section 1.2.2. La figure 5.13 présente l'environnement dans lequel ces évaluations ont eu lieu. Rappelons que l'objectif ici est d'effectuer deux tâches robotiques consistant (i) à guider l'utilisateur novice vers un produit du supermarché, (ii) à suivre l'utilisateur lors de ses déplacements dans les rayons. Ces différentes tâches doivent être réalisées dans un environnement public en respectant les contraintes d'interaction imposées, similaires à celles mise en place sur le robot Rackham, *i.e.* garder le contact visuel avec l'utilisateur et maintenir une distance sociale $d_{suivi} \in [2; 2.5]$m entre le robot et l'utilisateur.

Concrètement, lors de la tâche de guidage, l'utilisateur définit un produit à atteindre. Après planification du chemin vers ce produit, le robot commence sa mission et vérifie la présence de l'utilisateur durant l'exécution de la mission. Lorsque ce dernier s'éloigne du robot ($d_{suivi} > 2.5$m), le robot s'arrête et attend l'utilisateur. Lorsque l'utilisateur s'approche à nouveau du robot, ce dernier reprend son parcours initial. La figure 5.14(a) présente le résultat de l'exécution d'un telle tâche. L'utilisateur (ligne bleu) se déplace alors librement au sein de l'envrionnement alors que le robot (ligne rouge) suit une trajectoire prédéfinie allant de sa position initiale (zone $\#2$) vers un produit situé en zone $\#3$.

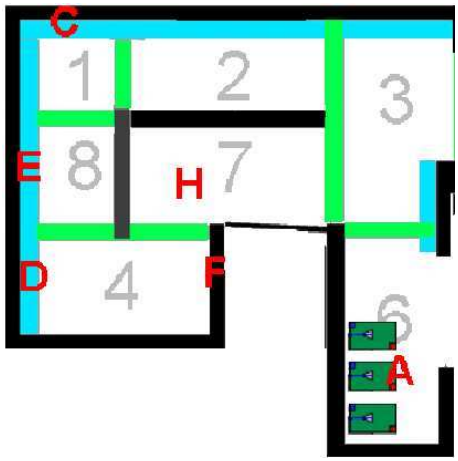


FIG. 5.13 – Description de l'environnement dans lequel évolue Inbot. Les zones délimitées par les lignes vertes et numérotées correspondent à des zones topologiques de l'environnement.

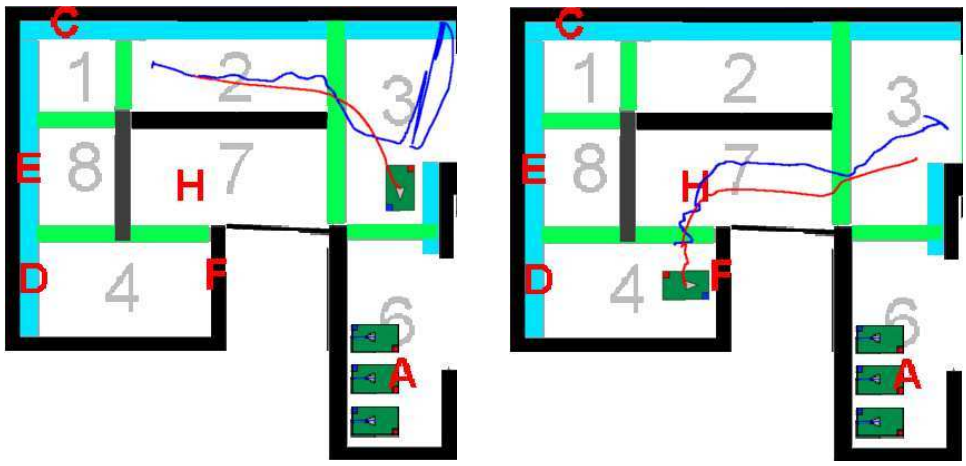
De façon similaire, lors d'une tâche de suivi, le robot réalise des mouvements conjointement à l'utilisateur afin de conserver le plus souvent possible ce dernier dans son champ de vision. Lorsque l'utilisateur s'éloigne du robot, ce dernier s'avance vers l'utilisateur afin de toujours conserver une distance sociale $d_{suivi} < 2.5$m. Lorsque le robot se trouve dans la zone d'interaction sociale, il s'arrête. Au contraire de Rackham, lors de cette tâche de suivi, le robot ne recule pas lorsque l'utilisateur s'approche. En effet, dans un contexte tel qu'un supermarché, l'utilisateur doit pouvoir accéder au panier du robot pour y déposer un article. La figure 5.14(b) montre le résultat de l'exécution d'un tâche de suivi. Le robot (ligne rouge) et l'utilisateur (ligne bleue) se déplacent de manière conjointe. L'utilisateur démarre en zone $\#3$ et se dirige vers le produit F situé en zone $\#4$.

Nous avons donc évalué les tâches robotiques dans leur ensemble, à savoir le ratio de missions de guidage et de suivi correctement effectuées par le robot sur le nombre total de missions.

En ce qui concerne les missions de guidage, la plupart des missions ont été réalisées avec succès. Parmi les échecs qui sont apparus, une faible partie est liée à un problème de planification ou de navigation (indépendant de notre système). La majeure partie correspond principalement à un problème de réinitialisation du suivi sur la base du RFID. En effet, l'environnement de type supermarché augmente considérablement le nombre de fausses détections RFID dues aux réflexions sur les structures métalliques environnantes.

Dans le cas des missions de suivi, 83% des missions ont été réalisées avec succès. De façon similaire, la plupart des échecs sont dûs aux fausses détections. Cependant, il est à noter que l'utilisateur se sert inconsciemment du badge RFID pour capter l'attention du robot, *i.e.* lorsque l'utilisateur s'aperçoit que le robot ne le suit plus, il 'montre' le badge RFID au robot qui s'oriente en direction de l'utilisateur.

La durée moyenne d'une mission a aussi été évaluée. Chaque mission, que ce soit une mission de guidage ou de suivi, met en jeu un parcours d'environ 10m. Lors d'une mission de guidage,



(a) (b)

FIG. 5.14 – Exécution des tâches robotiques de guidage (a) et de suivi (b). La ligne rouge (resp. bleue) représente la trajectoire du robot (resp. le trajectoire estimée de l'utilisateur).

la durée moyenne est d'environ $30s$ avec cependant un maximum de $55s$ pour les situations les plus difficiles (contre-jour, fausses détections RFID, ...). Pour ce qui est des missions d'accompagnement, leur durée moyenne est d'environ $45s$ avec un maximum de $65s$ dans des situations complexes. Bien que ces temps de parcours soient assez longs en situations réelles, ces résultats sont néanmoins encourageants car le robot doit prendre en compte les obstacles de l'environnement et les situations encombrées lors de ses déplacements qui, pour des raisons de sécurité, ont été limités.

5.4 Evitement de personnes : détection multi-personnes

Dans cette section, nous traitons du problème d'évitement de passants lors d'une mission de suivi, d'accompagnement ou de guidage, *i.e.* réaliser une manœuvre d'évitement tout en conservant l'utilisateur dans le champ de vue de la caméra. Dans ce but, nous utilisons les données issues des détecteurs de personnes définis en section 2.3. En effet, à ce jour, le suivi multi-cible décrit au chapitre 4 est en cours d'intégration sur notre plateforme. Cependant, les détecteurs laser et visuel de personnes sont assez robustes pour fournir une base satisfaisante à l'évaluation d'une tâche robotique d'évitement d'obstacles.

Nous allons donc, dans un premier temps, présenter la stratégie de contrôle utilisée avant de détailler les lois de commandes mises en œuvre. Bien que ce travail ne s'inscrive pas dans le cadre de cette thèse (travaux de Adrien Durand-Petiteville), sa présentation est indispensable à l'exécution et à l'évaluation de la tâche définie ci-dessus.

Par la suite, des évaluations, tant qualitatives que quantitatives, sont détaillées afin de valider la stratégie d'évitement de personnes basées sur une perception robot-centrée de l'environnement.

5.4.1 Loi de commande basée capteurs pour la tâche d'évitement de personnes

Dans le domaine de l'évitement d'obstacles, la recherche se divise traditionnellement en deux grandes catégories : l'approche globale et l'approche locale. Nous avons délibérément choisi d'utiliser une approche locale car aucune méthode de planification n'entre en compte dans nos travaux et les mouvements se doivent d'être les plus réactifs possibles, à la manière d'un reflexe chez l'homme.

Les méthodes locales que nous considérons ici utilisent en temps réel les informations perceptuelles qui renseignent sur l'environnement proche du robot. Sur la base de ces données, l'objectif est alors de synthétiser une commande qui guide le robot vers son but (*i.e.* l'utilisateur) tout en tenant compte de la présence d'obstacles. Les capteurs employés permettent alors de caractériser localement l'obstacle le plus proche du robot *i.e.* le risque de collision.

L'idée de la méthode des potentiels rotatifs [Folio, 2007] utilisée ici est de générer une force répulsive autour de l'obstacle qui impose au robot de suivre une enveloppe particulière, dite enveloppe de sécurité. Contrairement aux méthodes de potentiels classiques qui ne font que repousser le robot, le potentiel répulsif doit être capable de maintenir la commande du robot au voisinage de l'obstacle afin de pouvoir le contourner, et celà même en l'absence de potentiel attractif (*i.e.* absence de cible).

La méthode des potentiels rotatifs défini alors les vitesses linéaire et angulaire $(v_{coll}, \omega_{coll})$ nécessaire à l'évitement d'un obstacle.

Notre objectif est de synthétiser des lois de commande permettant à notre robot mobile de naviguer dans un environnement contraint sur la base des données visuelles perçues par la caméra. Pour celà, il est nécessaire d'intégrer les fonctionnalités d'évitement d'obstacles que nous venons de présenter succintement. La stratégie adoptée ici consiste à réaliser la tâche de suivi telle que définie à la section précédente lorsqu'il n'y a pas d'obstacles dans le voisinage du robot, et à basculer sur l'évitement d'obstacles lorsqu'un risque de collision se présente.

La méthode envisagée ici consiste à synthétiser séparément la commande d'évitement et celle guidant le robot vers l'utilisateur, puis à fusionner les deux correcteurs obtenus par une combinaison convexe. Il est ainsi possible de définir une loi de commande globale permettant de réaliser la tâche de navigation désirée telle que :

$$\begin{pmatrix} v_r^f \\ \omega_r^f \\ \omega_t^f \end{pmatrix} = (1 - \mu_{coll}) \begin{pmatrix} v_r \\ \omega_r \\ \omega_t \end{pmatrix} + \mu_{coll} \begin{pmatrix} v_{coll} \\ \omega_{coll} \\ 0 \end{pmatrix}, \tag{5.2}$$

où $\mu_{coll} \in [0; 1]$ est une fonction de la distance d_{coll} entre le robot et l'obstacle qui permet de lisser le basculement d'un correcteur à l'autre, et où $(v_r, \omega_r, \omega_t)'$ sont tels que définie dans l'équation 5.1.

De plus, afin de compenser les mouvements du robot lors de l'évitement d'une personne et de conserver l'utilisateur dans le champ de vue, il est nécessaire de contrôler l'orientation de la

platine indépendemment du robot. La vitesse angulaire en pan de la platine ω_p est alors définie par :

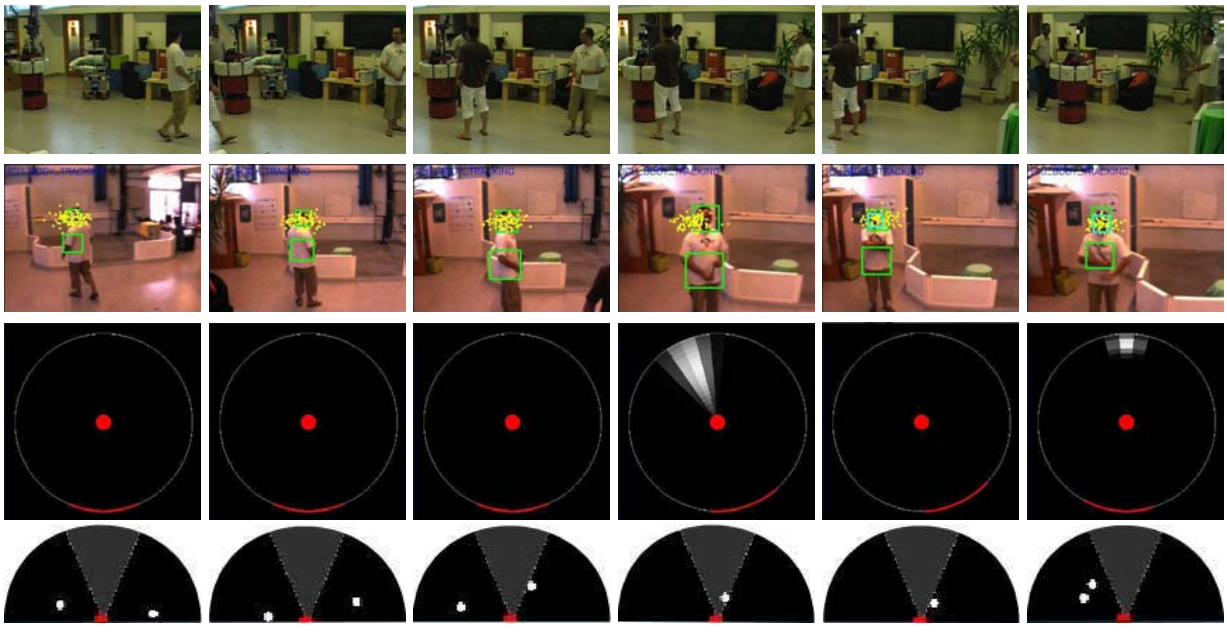
$$\omega_p = -\omega_r^f + \omega_r, \tag{5.3}$$

et correspond à la consigne initiale basée sur les données images (ω_r) auquel on ajoute le mouvement inverse du robot (ω_r^f).

5.4.2 Expérimentations sur Rackham et discussions associées

Une campagne d'évaluations préliminaires a été réalisée dans le même genre de configurations qu'à la section 5.3.2. Un niveau de complexité est ici ajouté. En effet, en plus d'un asservissement visuel sur l'utilisateur, le robot est placé en présence de passants et doit donc les éviter. Rappelons que notre objectif dans ces évaluations est d'effectuer une tâche consistant à accompagner une personne novice, équipée d'un badge RFID, dans un lieu dynamique, naturel et encombré tout en évitant les passants présents dans cet environnement.

La figure 5.15 détaille les images clés du flux vidéo ainsi que les sorties du traqueur, la carte de saillance RFID associée et les détections laser de personnes dans des conditions nominales.



(a) (b) (c) (d) (e) (f)

FIG. 5.15 – Images clés d'une séquence réalisée en conditions nominales. La première ligne représente la situation Homme / Robot. La deuxième ligne montre le flux vidéo et le résultat du suivi de l'utilisateur. La troisième ligne représente la carte de saillence RFID. La quatrième ligne est une vue robot-centrée des détections laser de personnes.

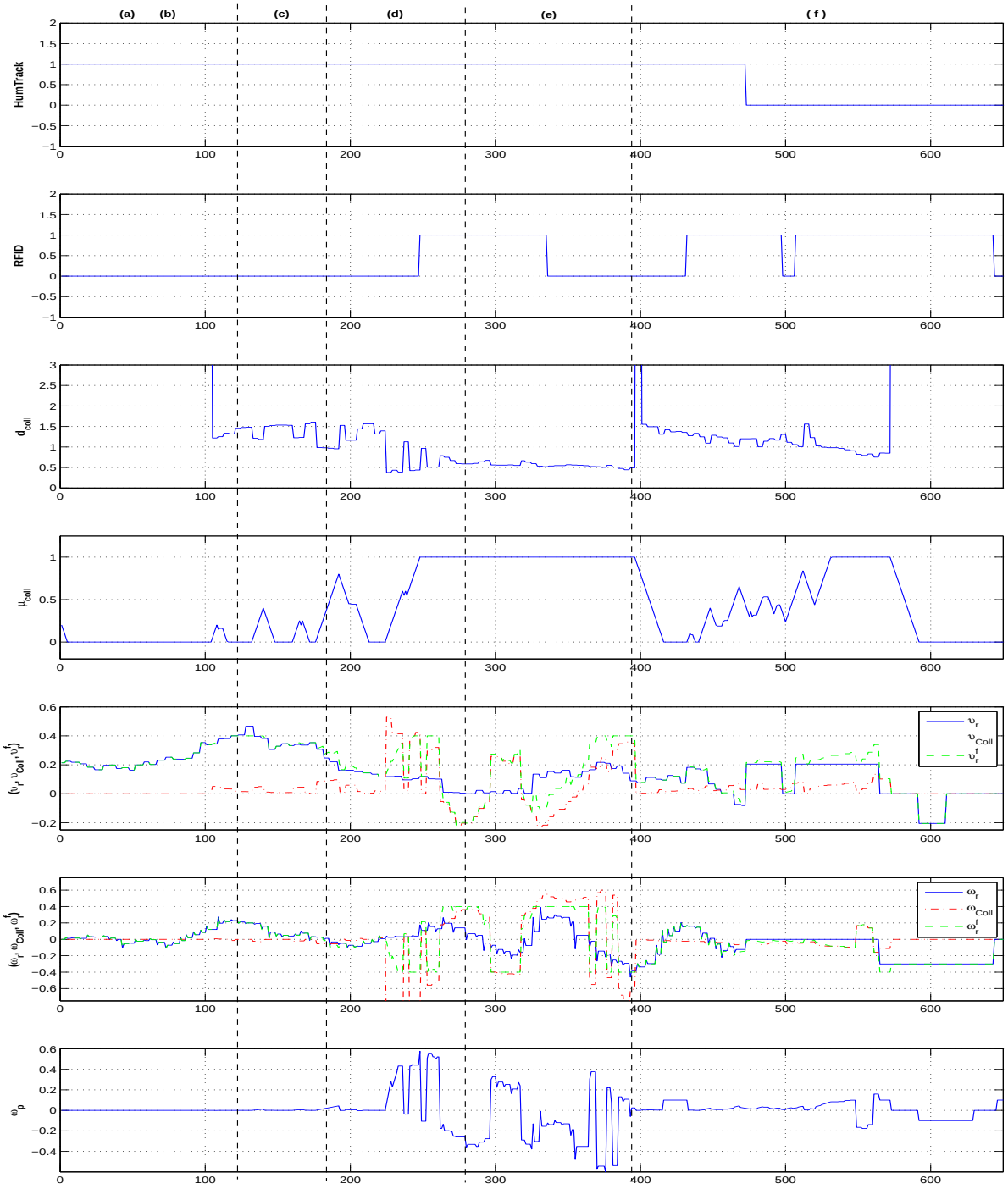


FIG. 5.16 – Synchronisation des flux de données entre les différents modules.

La figure 5.16 montre alors les différents signaux des modules `HumTrack`, `RFID`, `VisuServ` à savoir :

- les deux variables `HumTrack` et `RFID` qui prennent la valeur 1 lorsque la cible est détectée respectivement dans l'image et par le système RFID,
- les valeurs de d_{Coll} représentant la distance du robot à l'obstacle le plus proche et μ_{coll} permettant la fusion des vitesses du robot calculées sur la base des consignes image et obstacle,
- les trois valeurs (v_r, v_{Coll}, v_r^f) calculées par le module `VisuServ` relative à la vitesse linéaire du robot,
- les trois valeurs $(\omega_r, \omega_{Coll}, \omega_r^f)$ calculées par le module `VisuServ` relative à la vitesse angulaire du robot,
- la valeur ω_p calculée d'après l'équation 5.3.

Après initialisation de la mission, le robot focalise son attention sur la personne cible (a-b), tout comme lors d'une mission d'accompagnement. Les lois de commande sont alors entièrement déduites des données fournies par le traqueur car il n'y a pas d'obstacles à proximité du robot (*i.e.* $\mu_{Coll} = 0$). Lorsqu'un passant s'approche (c), il est détecté comme étant un obstacle. La valeur de μ_{Coll} croît alors proportionnellement à la distance à cet obstacle d_{Coll}. Les lois de commandes résultantes (v_r^f, ω_r^f) fusionnent alors les données déduites du traqueur (v_r, ω_r) avec celles issues de l'évitement d'obstacles $(v_{Coll}, \omega_{Coll})$. Au fur et à mesure que le robot s'approche d'un passant (d-e), les lois de commandes finales privilégient l'évitement de l'obstacle au détriment du suivi de l'utilisateur. Lors de l'exécution de cette manœuvre visant à éviter un passant (*i.e.* $\mu_{Coll} \neq 0$), la platine orientable est commandée par l'intermédiaire de ω_p de manière à compenser le mouvement du robot qui ne fait plus face à l'utilisateur. Une fois l'obstacle dépassé (*i.e.* $\mu_{Coll} = 0$) (f), la platine est repositionnée dans l'axe du robot et ce dernier compense automatiquement ce mouvement en s'orientant en direction de l'utilisateur.

Nous avons ensuite réalisé une série de tests afin d'évaluer de manière qualitative et quantitatives la bonne exécution d'une mission. Ces évaluations ont été effectuées par plusieurs personnes, jouant alternativement le rôle de l'utilistaur ou du passant. Tout comme précédemment, un trajet a été défini pour l'utilisateur alors que les passants se déplacent de manière libre. La figure 5.17 décrit la configuration typique de l'environnement et le trajet de l'utilisateur lors de telles évaluations. Plusieurs personnes sont suposées passer entre le robot et l'utilisateur, occulter l'utilisateur et perturber le déplacement du robot en restant sur sa trajectoire. La figure 5.18 montre une séquence type des tests effectués dans l'environnement décrit par la figure 5.17.

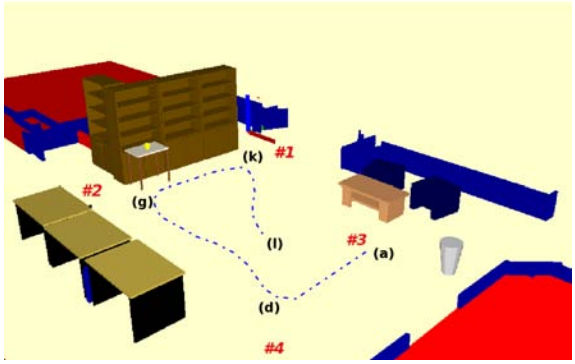


FIG. 5.17 – Exécution d'une tâche robotique d'accompagnement avec évitement d'obstacles.

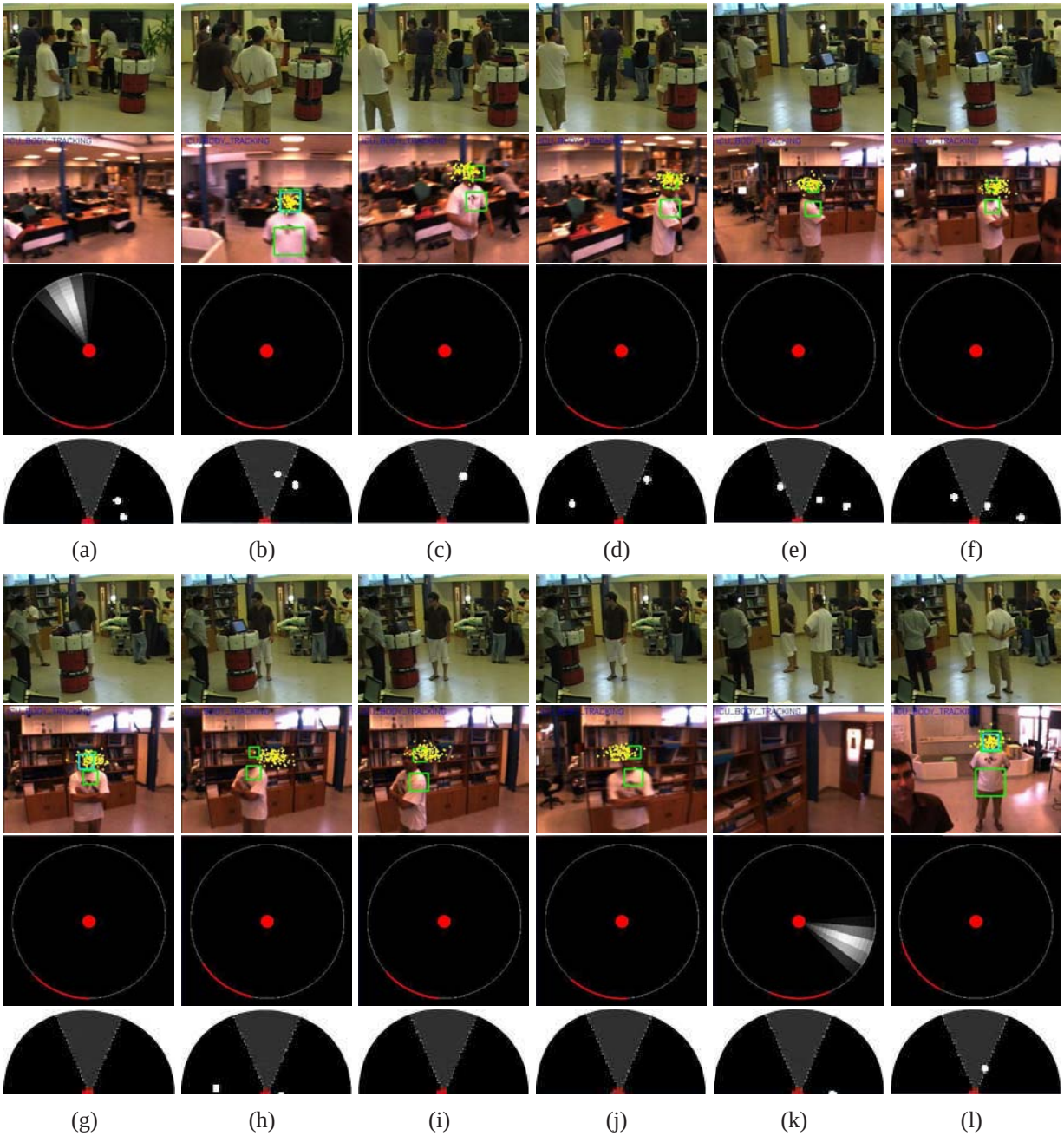


FIG. 5.18 – Exemple d'une séquence. La première ligne représente la situation Homme / Robot. La deuxième ligne montre le flux vidéo et le résultat du suivi de l'utilisateur. La troisième ligne représente la carte de saillence RFID. La quatrième ligne est une vue robot-centrée des détection laser de personnes.

Nous avons utilisés les mêmes métriques que celles utilisées précédemment, à savoir le *Ratio de Contact Visuel* et l'*Erreur de Guidage*, pour quantifier les performances de notre stratégie évaluée.

Sur l'ensemble des 10 réalisations effectuées, 8 d'entre elles se sont conclues par un succès. Les 2 échecs constatés sont dûs (1) à la présence d'un trop grand nombre de cibles autour du robot l'empêchant d'évoluer vers l'utilisateur, (2) à l'aspect sporadique des détections laser entraînant parfois des mouvements saccadés du robot.

En ce qui concerne le Ratio de Contact Visuel, l'utilisateur reste dans le champ de vision du robot plus de $\mu_{RCV} = 80\%$ du temps malgrè le couplage du système de suivi avec une stratégie d'évitement de personnes. Ceci est en grande partie dû à l'ajout du mouvement de compensation de la platine lors de la phase d'évitement. En effet, si une telle méthode n'est pas mise en place, le contact visuel avec l'utilisateur est perdu dès le début de la phase d'évitement d'un obstacle. Des évaluations complémentaires sur l'apport de cette heuristique sont en cours. Nous avons également mesuré l'erreur de guidage pendant ces tests. Sa valeur moyenne est de $\mu_{E_{suivi}} = 0.25$cm. L'augmentation de cette valeur comparée à celle du système n'intégrant pas l'évitement de personnes est dû au fait que, lors de la phase d'évitement, l'utilisateur continue d'évoluer dans l'environnement alors que le robot ralenti, impliquant une augmentation de la distance Homme / Robot.

Pour finir, il est à noter que, lors de l'ensemble des missions, le robot ne s'est jamais approché à plus de 50cm d'un passant. En effet, l'évolution de la valeur μ_{Coll} est telle que lorsque le robot est à 50cm d'un obstacle, la priorité exclusive est donnée à la phase d'évitement d'obstacle.

5.5 Conclusion

Nous avons, dans ce chapitre, présenté des évaluations de notre système de perception intégré sur différentes plateformes robotiques et dans différents contextes. Ces études nous ont permis de tester la robustesse de nos diverses fonctions perceptuelles aux contraintes de notre contexte applicatif et de nous fournir des résultats qualitatifs , mais aussi quantitatifs, permettant de prouver l'intérêt et la faisabilité de notre approche globale. En outre, l'association de la fusion de données hétérogènes pour le suivi et l'identification de l'utilisateur à un algorithme d'asservissement visuel donne des résultats très prometteurs quant à l'exécution d'une tâche interactive entre un homme et un robot. Il est à noter que ce travail conséquent d'intégration et d'évaluation robotique est peu présent dans la littérature.

Actuellement, des développements sont toujours en cours afin d'intégrer l'algorithme de suivi multi-cibles sur notre plateforme Rackham. En effet, nous avons constaté que les détecteurs de personnes, bien que très robustes, peuvent parfois 'oublier' une cible, entraînant une discontinuité dans la fonction perceptuelle et la loi de commande qui en résulte. L'utilisation du suivi permettrait alors de pallier à ce problème en lissant les données issues des détecteurs de personnes. De plus, l'analyse spatio-temporelle des trajectoires des différents passants permettrait

de mettre en place une fonction d'évitement d'obstacle plus élaborée afin d'éviter les situations bloquantes comme nous l'avons cité plus haut.

Conclusion

Dans ce manuscrit, nous avons présenté nos travaux visant à concevoir une interface perceptuelle multimodale de l'Homme depuis une plateforme mobile pour l'interaction Homme / Robot. Pour ce faire, nous avons détaillé un système complet de perception, depuis la mise en œuvre de capteurs embarqués, jusqu'à l'intégration et l'évaluation de tâches robotiques interactives, en passant par l'utilisation d'algorithmes de suivi et de fusion de données hétérogènes. Au delà des évaluations finales dans un contexte précis, chaque composante de la chaîne de perception, à savoir les détecteurs et identifieurs de personnes et les algorithmes de suivi, a été évaluée séparément afin de démontrer la robustesse de chaque brique qui compose l'architecture globale que nous avons proposée. L'ensemble des expérimentations qui ont été menées au cours de cette thèse en est la meilleure illustration. Nos travaux se doivent également d'être assez génériques pour faciliter leur intégration sur différentes plateformes robotiques sans besoin d'adaptation, comme celà a été le cas ici. Enfin, évoluant dans un contexte de robotique mobile autonome, nous respectons les contraintes propres à tout robot mobile autonome.

Plus concrètement, nos travaux visent à doter un robot assistant autonome de capacités lui permettant d'identifier son utilisateur, de conserver, tant que faire se peut, un contact visuel avec lui, de le suivre ou de le guider à travers un environnement dynamique, de percevoir les passants présents dans l'environnement et de les éviter.

Ce document débute par une introduction présentant le contexte et les objectifs de nos travaux. La problématique de la perception de l'homme par un robot assistant est introduite au travers d'un état de l'art des systèmes autonomes en interaction avec l'homme. Ensuite, le contexte spécifique de nos travaux, à savoir le projet CommRob, est présenté afin de cibler les points clés de notre approche et des scénarii permettant l'évaluation de cette approche sont détaillés. Enfin, nous décrivons notre approche ainsi que les principales spécificités qu'elle comporte.

Le deuxième chapitre traite de la détection et de la reconnaissance de l'homme au moyen de différents capteurs. Tout d'abord, une approche visuelle de la reconnaissance de visage est proposée. Cette approche est basée sur la construction d'une **base ACP globale** représentant l'ensemble des visages à reconnaître et la classification par **SVM multi-classes**. L'analyse des performances de ce classifieur au regard de la littérature est faite au moyen de **courbes ROC** et l'optimisation des paramètres libres du système utilise un **algorithme génétique NSGA-II**, assurant la sélection des paramètres optimaux en fonction du contexte et des contraintes temporelles fortes. Les résultats de classification sur des bases acquises depuis nos plateformes montrent l'intérêt d'une telle approche. Ensuite, l'adaptation d'un **système RFID passif** du marché est

proposée afin de pouvoir être embarqué sur une plateforme autonome. La réalisation d'une carte de multiplexage permet de répartir 8 antennes autour du robot afin de détecter et d'identifier le porteur d'un badge RFID sur $360°$ autour du robot. Des évaluations de robustesse à l'environnement sont détaillées et un modèle du capteur et proposé. Une étude visant à réduire la compacité globale du système est actuellement en cours. Pour finir, **deux détecteurs de personnes**, l'un basé sur des données **laser**, et l'autre basé **vision**, sont présentés. L'ensemble de ces fonctionnalités constituent une première couche d'abstraction dans notre architecture globale.

Le troisième chapitre présente une méthode de **fusion de données hétérogènes** pour le **suivi mono-cible**. L'utilisation d'un filtre particulaire de type $\mathrm{ICONDENSATION}$ facilite la fusion de données dans un cadre probabiliste. Nous proposons une fusion de données multi-capteurs basée sur la construction de **cartes de saillance** et un **algorithme d'échantillonnage par rejet** permettant d'échantillonner les particules du filtre de manière pertinente. Des évaluations sur la fusion de données au sein de la fonction d'importance du filtre, très peu abordée dans la littérature, sont proposées et permettent de confirmer l'apport indéniable (i) de la méthode d'échantillonnage en général et (ii) du système RFID en particulier, pour (re-)concentrer les particules au bon endroit de l'image et (ré-)initialiser le filtre.

Une **extension** sur l'utilisation de la fusion de données hétérogènes pour le **suivi multi-cible** est présentée au chapitre quatre. Les détections de personnes basées sur la vision et le laser, ainsi que les détections RFID sont utilisées comme données d'entrée d'un algorithme $\mathrm{RJ-MCMC}$ qui permet la gestion d'un **nombre de cibles variables** au cours du temps. Le principe de fusion de données détaillé au chapitre précédent est utilisé pour gérer l'entrée, la sortie ou la mise à jour d'une cible au sein de l'algorithme. Des évaluations préliminaires sont proposées.

Le cinquième et dernier chapitre décrit **l'intégration** de ces différentes fonctions perceptuelles sur **deux plateformes différentes**, Rackham et Inbot, et leur évaluation d'un point de vue robotique.

L'intégration sur nos plateforme robotiques étant, comme nous l'avons dit, un point clé de nos travaux, nous avons mené durant cette thèse des **expérimentations** dont le but a été de **valider** de manière incrémentale notre approche de la perception sur l'Homme, les couches d'abstraction qui la compose et leur symbiose dans cet ensemble. Ces démonstrations, concluantes sur plusieurs plateformes et dans différents cas d'utilisation, prouvent l'intérêt, la généricité et la validité de notre approche, tout en **mettant à l'épreuve** nos divers modules pour montrer leur robustesse. Ces évaluations robotiques, montrant l'intérêt de la perception de l'homme pour l'interaction Homme / Robot, ont fait l'objet du dernier chapitre de cette thèse.

L'ensemble de ces travaux ont donné lieu à 1 chapitre de livre, 3 articles de revue (dont un en cours de révision), 8 articles de conférences internationales (dont trois en collaboration) et 2 articles de conférence nationales.

Plus généralement, cette thèse a permis la combinaison de différentes techniques et différents domaines existants, *i.e.* un travail de synthèse, mais aussi un travail plus prospectif ouvrant de nombreuses voies d'applications pour la réalisation de tâches conjointes Homme / Robot. Elle débouche alors sur un certain nombre de perspectives à plus ou moins long terme.

Concernant directement nos travaux, plusieurs voies restent encore à explorer. Tout d'abord, il serait judicieux d'intégrer les détecteurs de personnes basés laser et vision dans la boucle de suivi mono-personne. En effet, il a été signalé que l'estimation de la distance Homme / Robot est peu précise. Or, les détections laser, par exemple, permettent de connaître la distance entre le robot et une cible donnée. L'estimation de la distance Homme / Robot en serait alors améliorée.

De plus, il serait intéressant d'étendre le modèle de l'utilisateur en y ajoutant, par exemple, des points d'intérêt, en complément de la mesure colorimétrique. Ce modèle pourrait alors être mis à jour et enrichi avec de nouvelles informations tout au long de la mission.

Des évaluations plus exhaustives sont en cours sur l'algorithme de suivi multi-personnes, en même temps qu'un portage sur notre plateforme Rackham. Il est ensuite nécessaire de quantifier l'apport de cette fonction de suivi par rapport aux détecteurs utilisés ici. Un élément de réponse porte à penser que l'utilisation d'un flux continu relatif aux obstacles présents dans l'environnement constituerait un bon support pour la fusion des lois de commandes relatives au suivi de personne et à l'évitement d'obstacles. De plus, la méthode pourrait facilement être étendu à l'ensemble des obstacles de l'environnement, et non plus seulement les personnes.

A plus long terme, il paraît intéressant de jouer sur la complémentarité entre les capteurs embarqués et déportés. En effet, de plus en plus d'approches instrumentent l'environnement afin d'obtenir une analyse plus fine de ce dernier. L'utilisation de caméras déportées dans l'environnement, actives ou omnidirectionnelles, est de plus en plus fréquent et peut apporter une richesse d'information sur l'environnement, ne serait-ce que par la puissance de calcul, qu'un robot seul ne peut analyser.

Pour finir, il est nécessaire d'évaluer la robustesse et la pertinence des tâches robotiques dans un cadre plus ouvert, c'est à dire avec des utilisateurs novices, afin de pouvoir juger de l'acceptabilité du robot assistant lui-même dans un environnement humain très encombré. En effet, ce type d'évaluations 'grand-public' reste encore peu abordé bien qu'indispensable pour l'introduction définitive des robots dans la vie de tous les jours.

Liste des publications

Revues et chapitres

Germa, T., Lerasle, F., Ouadah, N., and Cadenat, V. (2010). Vision and rfid data fusion for tracking people in crowds by a mobile robot. In *Computer Vision and Image Understanding, Special Issue on Multi-camera and Multi-modal Sensor Fusion (CVIU'10)*, volume 114, pages 641–651.

Germa, T., Lerasle, F., and Simon, T. (2009). Video-based face recognition and tracking from a robot companion. In *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI'09)*, volume 23, pages 591–616.

Menezes, P., Lerasle, F., Diaz, J., and Germa, T. (2007). *Towards an interactive humanoid companion with visual tracking modalities.*

Ouadah, N., Cadenat, V., Lerasle, F., Hamerlain, M., Germa, T., and Boudjema, F. (2010). A multi-sensor-based control strategy for initiating and maintaining interaction between a robot and a human. In *Journal of Advanced Robotics (submitted to)*.

Conférences internationales

Fontmarty, M., Germa, T., Burger, B., Marin, L.-F., and Knoop, S. (2007). Implementation of human perception algorithms on a mobile robot. In *6th IFAC Symposium on Intelligent Autonomous Vehicles (IAV'07)*, Toulouse, France.

Germa, T., Brèthes, L., Lerasle, F., and Simon, T. (2007a). Data fusion and eigenface based tracking dedicated to a tour-guide robot. In *Int. Conf. on Vision Systems (ICVS'07)*, Bielefeld, Germany.

Germa, T., Devy, M., Rioux, R., and Lerasle, F. (2009a). A tuning strategy for face recognition in robotic application. In *Int. Conf. on Computer Vision Theory and Applications (VISAPP'09)*, Lisbon, Portugal.

Germa, T., Lerasle, F., Danès, P., and Brèthes, L. (2007b). Human/robot visual interaction for a tour-guide robot. In *Int. Conf. on Intelligent Robots and Systems (IROS'07)*, San Diego, USA.

Germa, T., Lerasle, F., Ouadah, N., and Cadenat, V. (2009b). Vision and rfid-based person tracking in crowds from a mobile robot. In *22nd IEEE Int. Conf. on Intelligent Robots and Systems (IROS'09)*, St Louis, USA.

Göller, M., Devy, M., Steinhardt, F., Germa, T., Lerasle, F., Kerscher, T., and Dillmann, R. (2010a). Control-sharing and trading of a service robot acting as intelligent shopping cart. In *19th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'10)*, Viareggio, Italy.

Göller, M., Kerscher, T., Zollner, J., Dillmann, R., Devy, M., Germa, T., and Lerasle, F. (2009). Setup and control architecture for an interactive shopping cart in human all day environments. In *Advanced Robotics, 2009. ICAR 2009. International Conference on*, pages 1 –6.

Göller, M., Steinhardt, F., Szép, A., Ertl, D., Germa, T., , and Devy, M. (2010b). Modes of interaction for a cognitive shopping cart. In *4th International Conference on Cognitive Systems (CogSys'10)*, Zurich, Switzerland.

Conférences nationales

Germa, T., Brèthes, L., Lerasle, F., and Simon, T. (2007). Suivi et identification de personnes par un robot guide. In *11ème congrès francophone des jeunes chercheurs en vision par ordinateur (ORASIS'07)*, Obernai, France.

Germa, T., Lerasle, F., Ouadah, N., Cadenat, V., and Lemaire, C. (2010). Fusion de données visuelles et rfid pour le suivi de personnes en environnement encombré depuis un robot mobile. In *17e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA'10)*, Caen, FRANCE.

Bibliographie

Abata, A., Nappi, M., Riccio, D., and Sabatino, G. (2007). 2D and 3D face recognition : a survey. *Pattern Recognition Letters*, 28(14) :1885–1906.

Adini, Y., Moses, Y., and Ullman, S. (1997). Face recognition : the problem of compensating for changes in illumination direction. *Trans. on Pattern Analysis Machine Intelligence (PAMI'97)*, 19(7) :721–732.

Aggarwal, G., Roy-Chowdhury, A., and Chepalla, R. (2004). A system identification approach for video-based face recognition. In *Int. Conf. on Pattern Recognition (ICPR'04)*, Cambridge, UK.

Aherne, F., Thacker, N., and Rockett, P. (1997). The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 32(4) :1–7.

Alami, R., Chatila, R., Fleury, S., and Ingrand, F. (1998). An architecture for autonomy. *Int. Journal of Robotic Research (IJRR'98)*, 17(4) :315–337.

Anne, M., Crowley, J., Devin, V., and Privat, G. (2005). Localisation intra-bâtiment multi-technologies : RFID, wifi et vision. In *National Conf. on Mobility and ubiquity computing (UbiMob'05)*, pages 29–35, Grenoble, France.

Arulampalam, S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *Trans. on Signal Processing*, 2(50) :174–188.

Azimi-Sadjadi, B. and Krishnaprasad, P. S. (2004). A particle filtering approach to change detection for nonlinear systems. *EURASIP J. Appl. Signal Process.*, 2004 :2295–2305.

Bardet, F. and Chateau, T. (2008). Mcmc particle filter for real-time visual tracking of vehicles. In *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, pages 539 –544.

Bartlett, M., Movellan, J., and Sejnowski, T. (2002). Face recognition by independent component analysis. *Trans. on Neural Networks*, 13(6) :1450–1464.

Belhumeur, P., Hespanha, J., and Kriegman, D. (1996). Eigenfaces vs. fisherfaces. In *European Conf. on Computer Vision (ECCV'96)*, pages 45–58.

Bellotto, N. and Hu, H. (2006). Vision and laser data fusion for tracking people with a mobile robot. In *Int. Conf. on Robotics and Biomimetics (ICRB'06)*, Kunming, China.

Bennewitz, M., Faber, F., Joho, D., Schreiber, M., and Behnke, S. (2005). Towards a humanoid museum guide robot that interacts with multiple persons. In *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, pages 418–423.

Bischoff, R. and Graefe, V. (2004). Hermes - a versatile personal robotic assistant. In *Human Interactive Robots, Proceedings of IEEE*, pages 1759–1779.

Bischoff, R., Kazi, A., and Seyfarth, M. (2002). The morpha style guide for icon-based programming. In *Robot and Human Interactive Communication, 2002 (ROMAN'02). Proceedings. 11th IEEE International Workshop on*, pages 482 – 487, Berlin, Germany.

Boardman, M. and Trappenberg, T. (2006). A heuristic for free parameter optimization with SVM. In *Int. Joint Conf. on Neural Networks (IJCNN'06)*, pages 610–617, Pula, Croatie.

Bohme, H., Wilhelm, T., Key, J., Schauer, C., Schroter, C., H.M., G., and T., H. (2003). An approch to multi-modal human-machine interaction for intelligent service robot. 44 :83–96.

Bonnal, J., Argentieri, S., Danès, P., and Manhès, J. (2009). Speaker localization and speech extraction with the ear sensor. In *22nd IEEE Int. Conf. on Intelligent Robots and Systems (IROS'09)*.

Bregonzio, M., Taj, M., and Cavallaro, A. (2007). Multimodal particle filtering tracking using appearance, motion and audio likelihoods. In *Int. Conf. on Image Processing (ICIP'07)*, San Antonio, USA.

Brèthes, L. (2005). *Suivi visuel par filtrage particulaire. Application à l'interaction Homme-Robot*. PhD thesis, Université Paul Sabatier de Toulouse.

Burgard, W., Fox, D., Lakemeyer, G., Haehnel, D., Schulz, D., Steiner, W., Thrun, S., and Cremers, A. (1998). Real robots for the real world — the rhino museum tour-guide project. In *Proceedings of the 1998 AAAI Spring Symposium*.

Burger, B. (2010). *Fusion de données audio-visuelles pour l'interaction Homme-Robot*. PhD thesis, Université Paul Sabatier de Toulouse.

Calisi, D., Iocchi, L., and Leone, R. (2007). Person following through appearance models and stereo vision using a mobie robot. In *Int. Conf. on Computer Vision Theory and Applications (VISAPP'07)*, Barcelona, Spain.

Castano, B. and Rodriguez, M. (2008). An artificial intelligence and RFID system for people detection and orientation in big surfaces. In *Int. Multi-Conf. on Engineering and Technological Innovation (IMETI'08)*, Orlando, USA.

Chae, H. and Han, K. (2005). Combination of rfid and vision for mobile robot localization. pages 75 – 80.

Chapelle, O., Vapnik, V., Bousquet, O., and Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. *Machine Learning*, 46(1) :131–159.

Chateau, T., Goyat, Y., and Trassoudaine, L. (2009). M2sir : A multi modal sequential importance resampling algorithm for particle filters. pages 4073 –4076, Cairo, Egypt.

Chella, Antonio, Liotta, Marilia, Macaluso, and Irene (2007). Cicerobot : a cognitive robot for interactive museum tours. volume 34, pages 503–511. Emerald Group Publishing Limited.

Chen, C.-H. and Chan, Y.-P. (2007). Real time multi-target visual tracking based on velocity segmentation technique. In *Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE*, pages 2813 –2817.

Chen, L. L., Shih, T. K., and Hung, J. C. (2009). Using rfid to realize human computer interaction. pages 859 –864.

Chen, P., Lin, C., and Scholkopf, B. (2005). A tutorial on v-Support Vectors Machines. 21(2) :111–136.

Cho, J., Hun Jin, S., Dai Pham, X., and Jeon, J. (2007). Multiple object tracking circuit using particle filters with multiple features. In *Int. Conf. on Robotics and Automation (ICRA'07)*, pages 4639–4644, Roma, Italy.

Choudhury, T., Clarkson, B., Jebara, T., and Pentland, A. (1999). Multimodal person recognition using unconstrained audio and video. In *Int. Conf. on Audio- and Video-based Person Authentification*, pages 176–180.

Cielniak, G., Lilienthal, A., and Duckett, T. (2007). Improved data association and occlusion handling for vision-based people tracking by mobile robots. In *Int. Conf. on Intelligent Robots and Systems (IROS'07)*, San Diego, USA.

CLEAR (2007). *2007 Benchmark Data*. Online, Baltimore, MD, USA.

Clodic, A., Montreuil, V., Alami, R., and Chatila, R. (2005). A decisional framework for autonomous robots interacting with humans. In *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN'05)*.

Corke, P. (1996). *Visual control of robots : High performance visual servoing*. Research Studies Press LTD.

Cui, J., Zha, H., Zhao, H., and Shibasaki, R. (2008). Multimodal tracking of people using laser scanners and video camera. *Image and Vision Computing (IVC'08)*, 26(2) :240–252.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005 (CVPR'05). IEEE Conference on*, pages 886–893, San Diego, CA, USA.

Doucet, A., De Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Series Statistics For Engineering and Information Science. Springer-Verlag, New York.

Doucet, A., Godsill, S. J., and Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3) :197–208.

Durand-Petiteville, A., Cadenat, V., and Courdesses, M. (2010). A unified initialization phase to improve visual servoing in an unknown environment. In *7th Symposium on Intelligent Autonomous Vehicles (IAV'10)*, Lecce, Italie.

Espiau, B., Chaumette, F., and Rives, P. (1992). A new approach to visual servoing in robotics. 8 :313–326.

Ess, A., Leibe, B., Schindler, K., and Van Gool, L. (2008). A mobile vision system for robust multi-person tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http ://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. The PASCAL Visual Object Classes Challenge 2008 (VOC2008) Results. http ://www.pascal-network.org/challenges/VOC/voc2008/workshop/index.html.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2009). Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(PrePrints).

Folio, D. (2007). *Stratégies de commande référencées multi-capteurs et gestion de la perte du signal visuel pour la navigation d'un robot mobile*. PhD thesis, Université Paul Sabatier, LAAS-CNRS, LAAS-CNRS, Toulouse, FRANCE.

Gabriel, P., Verly, J., Piater, J., and Genon, A. (2003). The state of the art in multiple object tracking under occlusion in video sequences. In *Int. Conf. on Advanced Concepts for Intelligent Vision Systems (ACIVS'03)*, Ghent, Belgium.

Gavrila, D. and Munder, S. (2007). Multi-cue pedestrian detection and tracking from a moving vehicle. *Int. Journal of Computer Vision (IJCV'07)*, 73(1) :41–59.

Germa, T., Brèthes, L., Lerasle, F., and Simon, T. (2007a). Data fusion and eigenface based tracking dedicated to a tour-guide robot. In *Int. Conf. on Vision Systems (ICVS'07)*, Bielefeld, Germany.

Germa, T., Lerasle, F., Danès, P., and Brèthes, L. (2007b). Human/robot visual interaction for a tour-guide robot. In *Int. Conf. on Intelligent Robots and Systems (IROS'07)*, San Diego, USA.

Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2) :107–113.

Gorostiza, J., Barber, R., Khamis, A., and Malfaz, M. (2006). Multimodal human-robot framework for a personal robot. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'06)*, pages 39–44, Hatfield, UK.

Guo, G., Li, S., and Chan, K. (2000). Face recognition by support vector machines. In *Int. Conf. on Face and Gesture Recognition (FGR'00)*, pages 196–201, Grenoble, France.

Hahnel, D., Burgard, W., Fox, D., Fishkin, K., and Philipose, M. (2004). Mapping and localization with RFID technology. *Int. Conf. on Robotics and Automation (ICRA'04)*, pages 1015–1020.

Hall, E. (1966). *The Hidden Dimension*. Doubleday, Garden City, N.Y.

Hammoud, R. and Davis, J. (2007). Advances in vision algorithms and systems beyond the visible spectrum. *Computer Vision and Image Understanding (CVIU'07)*, 106(2) :145–147.

Harte, E. and Jarvis, R. (2007). Multimodal human-robot interaction in an assistive technology context. Brisbane, Australia.

Heisele, B., Ho, P., and Poggio, T. (2001). Face recognition with support vector machines. In *Int. Conf. on Computer Vision (ICCV'01)*, pages 688–694.

Heseltine, T., Pears, N., and Austin, J. (2002). Evaluation of image pre-processing techniques for eigenface based recognition. In *SPIE : Image and Graphics*, pages 677–685.

Huang, C., Ai, H., Li, Y., and Lao, S. (2007). High-performance rotation invariant multi-view face detection. *Trans. on Pattern Analysis Machine Intelligence (PAMI'07)*, 29(4) :671–686.

Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density. In *European Conf. on Computer Vision (ECCV'96)*, pages 343–356, Cambridge, UK.

Isard, M. and Blake, A. (1998a). CONDENSATION – conditional density propagation for visual tracking. *Int. Journal on Computer Vision*, 29(1) :5–28.

Isard, M. and Blake, A. (1998b). I-CONDENSATION : Unifying low-level and high-level tracking in a stochastic framework. In *European Conf. on Computer Vision (ECCV'98)*, pages 893–908.

Isard, M. and MacCormick, J. (2001). Bramble : a bayesian multiple-blob tracker. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 34 –41 vol.2.

Jia, S., Shang, E., Abe, T., and Takase, K. (2006). Localization of mobile robot with rfid technology and stereo vision. pages 508 –513.

Jia, S., Sheng, J., and Takase, K. (2008). Human recognition using rfid system with multi-antenna. pages 1213 –1218.

Jin, Y. (2009). Beyond icondensation : Aicondensation and afcondensation for visual tracking with low-level and high-level cues. pages 4089 –4092, Cairo, Egypt.

João, X., Pacheco, M., Castro, D., Ruano, A., and Nunes, U. (2005). Fast line, arc/circle and leg detection from laser scan data in a player driver. In *Int. Conf. on Robotics and Automation (ICRA'05)*, Barcelona, Spain.

Jonsson, K., Matas, J., Kittler, J., and Li, Y. (2000). Learning support vectors for face verification and recognition. In *Int. Conf. on Face and Gesture Recognition (FGR'00)*, pages 208–213, Grenoble, France.

Kanda, T., Ishiguro, H., Imai, M., and Ono, T. (2004). Development and evaluation of interactive humanoid robots.

Kanda, T., Shiomi, M., Perrin, L., Nomura, T., Ishiguro, H., and Hagita, N. (2007). Analysis of people trajectories with ubiquitous sensors in a science museum. In *Int. Conf. on Robotics and Automation (ICRA'07)*, pages 4846–4853, Roma, Italy.

Kar, B., Bhatia, S., and Dutta, P. (2007). Audio-visual biometric based speaker identification. In *Int. Conf. on Computational Intelligence and Multimedia Applications (ICCIMA'07)*, pages 94–98, Sivakasi, India.

Khan, Z., Balch, T., and Dellaert, F. (2005). MCMC-based particle filtering for tracking a variable number of interacting targets. *Trans. on Pattern Analysis Machine Intelligence (PAMI'05)*, 27(11) :1805–1818.

Kobilarov, M., Sukhatme, G., Hyams, J., and Batavia, P. (2006). People tracking and following with mobile robot using an omnidirectional camera and laser. In *Int. Conf. on Robotics and Automation (ICRA'06)*, pages 557–562, Orlando, USA.

Koch, J., Wettach, J., Bloch, E., and Berns, K. (2007). Indoor localisation of humans, objects, and mobile robots with rfid infrastructure. pages 271 –276.

Kubitz, O., Berger, M., Perlick, M., and Dumoulin, R. (1997). Application of radio frequency identification devices to support navigation of autonomous mobile robots. volume 1, pages 126 –130 vol.1.

Kulyukin, V., Gharpure, C., Nicholson, J., and Pavithran, S. (2004). Rfid in robot-assisted indoor navigation for the visually impaired. volume 2, pages 1979 – 1984 vol.2.

Kurazume, R., Yamada, H., Murakami, K., Iwashita, Y., and Hasegawa, T. (2008). Target tracking using sir and mcmc particle filters by multiple cameras and laser range finders. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 3838 –3844.

Lam, K. and Yan, H. (98). An analytic-to-holistic approach fo face recognition based on a single frontal view. *Trans. on Pattern Analysis Machine Intelligence (PAMI'98)*, 7(20) :673–686.

Lee, K., Ho, J., Yang, M., and Kriegman, D. (2003). Video-based face recognition using probabilistic appearance manifolds. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 1 :I–313–I–320 vol.1.

Li, Y., Ai, H., Huang, C., and Lao, S. (2006). Robust head tracking with particles based on multiple cues fusion. In *Computer Vision in Human-Computer Interaction (CVHCI06)*, pages 29–39.

Lieckfeldt, D., You, J., and Timmermann, D. (2009). Exploiting rf-scatter : Human localization with bistatic passive uhf rfid-systems. pages 179 –184.

Lin, K., K.M., L., and W., S. (2003). Spatially eigen-weighted Hausdorff distances for human face recognition. *Pattern Recognition (PR'03)*, 36(8) :1827–1834.

Liu, X., Corner, M. D., and Shenoy, P. (2006). Ferret : Rfid localization for pervasive multimedia. In *Proceedings of the 8th International Conference on Ubiquitous Computing*.

Maas, J., Spexard, T., Fritsch, J., Wrede, B., and Sagerer, G. (2006). Biron, what's the topic ? a multi-modal topic tracker for improved human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 26–32.

MIT (2000). *MIT Face Recognition Database*. http ://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html.

Mori, S. and Chong, C.-Y. (2008). Markov chain monte carlo method for evaluating multi-frame data association hypotheses. In *Information Fusion, 2008 11th International Conference on*, pages 1 –8.

Mori, T., Suemasu, Y., Noguchi, H., and Sato, T. (2004). Multiple people tracking by integrating distributed floor pressure sensors and RFID system. In *Int. Conf. on Systems, Man and Cybernetics*, pages 5271–5278, The Hague, Netherlands.

Muñoz Salinas, R., García-Silvente, M., and Medina-Carnicer, R. (2008). Adaptive multi-modal stereo people tracking without background modelling. *J. Vis. Comun. Image Represent.*, 19(2) :75–91.

Ni, L. M., Liu, Y., Lau, Y. C., and Patil, A. P. (2004). Landmarc : Indoor location sensing using active rfid. *Wireless Networks*, 10(6) :701–710.

Nickel, K., Gehrig, T., Ekenel, H., Stiefelhagen, R., and McDonough, J. (2005). A joint particle filter for audio-visual speaker tracking. In *Int. Conf. on Multimodal Interfaces (ICMI'05)*, pages 61–68, Torento, Italy.

Nourbakhsh, I., Kunz, C., and Willeke, T. (2003). The mobot museum robot installations : a five year experiment. In *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 4, pages 3636–3641 vol.3.

Nummiaro, K., Koller-Meier, E., and Gool, L. V. (2003). An adaptative color-based particle filter. *Image and Vision Computing (IVC'03)*, 21(90) :90–110.

Ouadah, N., Cadenat, V., F., B., and Hamerlain, M. (2009). Image based robust visual servoing on human face to improve human/robot interaction. In *4th European Conference on Mobile Robots (ECMR'09)*, pages 155–160, Mlini/Dubrovnick, Croatia.

Pang, Y., Liu, Z., and Yu, N. (2006). A new nonlinear extraction method for face recognition. *Neurocomputing*, (69) :949–953.

Pérez, P., Vermaak, J., and Blake, A. (2004). Data fusion for visual tracking with particles. *Proc. IEEE*, 92(3) :495–513.

Pérez, P., Vermaak, J., and Gangnet, M. (2002). Color-based probabilistic tracking. In *European Conf. on Computer Vision (ECCV'02)*, pages 661–675, Berlin.

PETS (2004). *2004 Benchmark Data*. Online, Prague, Czech Republic.

PETS (2006). *2006 Benchmark Data*. Online, New York, USA.

Phillips, P., Moon, H., Rizvi, S., and Rauss, P. (2000). The feret evaluation methodology for face-recognition algorithms. *Trans. on Pattern Analysis Machine Intelligence (PAMI'00)*, 22(10) :1090–1104.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., and Thrun, S. (2003). Towards robotic assistants in nursing homes : challenges and results. 42 :271–281.

Provost, F. and Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, 42(3) :203–231.

Qu, W., Schonfeld, D., and Mohamed, M. (2007). Distributed bayesian multiple-target tracking in crowded environments using multiple collaborative cameras. *EURASIP Journal on Advances in Signal Processing*.

Quintiliano, P., Santa-Rosa, A., and Guadagnin, R. (2001). Face recognition based on eigenfeatures. In *SPIE : Image extraction, segmentation and recognition*, pages 140–145.

Ryu, H. R. and Huber, M. (2007). A particle filter approach for multi-target tracking. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 2753 –2760.

Schroeter, C., Hoechemer, M., Mueller, S., and Gross, H.-M. (2009). Autonomous robot cameraman - observation pose optimization for a mobile service robot in indoor living space. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 424 –429.

Schulz, D., Fox, D., and Hightower, J. (2003). People tracking with anonymous and ID-sensors using rao-blackwellised particle filters. In *Int. Joint Conf. on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico.

Seo, K. (2007). A GA-based feature subset selection and parameter optimization of SVM for content-based image retrieval. In *Int. Conf. on Advanced Data Mining and Applications (ADMA'07)*, pages 594–604, Harbin, China.

Shan, S., Gao, W., and Zhao, D. (2003). Face recognition based on face-specific subspace. *Int. Journal of Imaging Systems and Technology*, 13(1) :23–32.

Siegwart, R., Arras, K. O., Bouabdallah, S., Burnier, D., Froidevaux, G., Greppin, X., Jensen, B., Lorotte, A., Mayor, L., Meisser, M., Philippsen, R., Piguet, R., Ramel, G., Terrien, G., and Tomatis, N. (2003). Robox at expo.02 : A large-scale installation of personal robots. *Robotics and Autonomous Systems*, 42(3-4) :203 – 222.

Smith, K., Gatica-Perez, D., and Odobez, J. (2005). Using particles to track varying numbers of interacting people. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, pages 962–969, San Diego, USA.

Spall, J. (2002). Estimation via markov chain monte carlo. In *American Control Conference, 2002. Proceedings of the 2002*, volume 4, pages 2559 – 2564 vol.4.

Spinello, L., Triebel, R., and Siegwart, R. (2008). Multimodal detection and tracking of pedestrians in urban environments with explicit ground plane extraction. In *AAAI Conf. on Artificial Intelligence (AAAI'08)*, pages 1409–1414, Chicago, USA.

Takahashi, S., Wong, J., and Miyamae, M. (2008). A ZigBee-based sensor node for tracking people's locations. In *ACM Int. Conf.*, pages 34–38, Sydney, Australia.

Thrun, S., Beetz, M., Bennewitz, M., Burgard, W., Cremers, A. B., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C., Roy, N., Schulte, J., and Schulz, D. (2000). Probabilistic algorithms and the interactive museum tour-guide robot minerva. *International Journal of Robotics Research*, 19(11) :972–999.

Trahanias, P., Argyros, A., Tsakiris, D., Cremers, A., Schulz, D., Burgard, W., Haehnel, D., Savvaides, V., Giannoulis, G., Coliou, Y., Kamarinos, G., Friess, P., Konstantios, D., and Katselaki, A. (2000). Tourbot - interactive museum tele-presence through robotic avatars. *Cultivate Interactive, issue*, 7.

Tsai, Y., Shih, H., and Huang, C. (2006). Multiple human objects tracking in crowded scenes. In *Int. Conf. on Pattern Recognition (ICPR'06)*, pages 51–54, Hong Kong.

Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'91)*, pages 586–591.

Viola, P. and Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'01)*.

Viola, P. and Jones, M. (2003). Fast multi-view face detection. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'03)*.

Waagepetersen, R. and Sorensen, D. (2001). A tutorial on reversible jump mcmc with a view toward applications in qtl-mapping. In *on QTL mapping. International Statistical Review 69, 49 - 62*, pages 200 – 1.

Wang, Y., Liu, Y., Tao, L., and Xu, G. (2006). Real-time multi-view face detection and pose estimation in video stream. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 4 :354–357.

Wu, T., Lin, C., and Weng, R. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5 :975–1005.

Xu, L. and Li, C. (2006). Multi-objective parameters selection for SVM classification using NSGA-II. In *Industrial Conference on Data Mining (ICDM'06)*, pages 365–376.

YALE (2006). *Yale Face Database*. http ://cvc.yale.edu/projects/yalefacesB/yalefacesB.html.

Yang, B., Pan, X., Men, A., and Chen, X. (2010). A robust particle filter for people tracking. *Future Networks, International Conference on*, 0 :20–23.

Yang, H., Jiao, X., Zhang, L., and Li, F. (2006). Parameter optimization for SVM using sequential number theoretic for optimization. In *Int. Conf. on Machine Learning and Cybernetics*, Dalian.

Yao, J. and Odobez, J. (2008). Multi-camera multi-person 3d space tracking with mcmc in surveillance scenarios. In *European Conference on Computer Vision, 2008 (ECCV'08). Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, Marseille, France.

Ying, L., Narayanan, S., and Kuo, C. (2004). Adaptative speaker identification with audiovisual cues for movie content analysis. *Pattern Recognition Letters*, 25(7) :776–791.

Zajdel, W., Zivkovic, Z., and Kröse, B. (2005). Keeping track of humans : have I seen this person before ? In *Int. Conf. on Robotics and Automation (ICRA'05)*, pages 2093–2098, Barcelona, Spain.

Zhao, T. and Nevatia, R. (2004). Tracking multiple humans in crowded environment. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–406 – II–413 Vol.2.

Zhao, W. and Chellapa, R. (1998). Discriminant analysis of principal components for face recognition. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'98)*, pages 336–341, Santa Barbara, USA.

Zhao, W., Chellappa, R., Phillips, P., and Rosenfeld, A. (2000). Face recognition : a literature survey. *ACM Computing Surveys*, 35(4) :399–458.

Zhou, Y., Liu, W., and Huang, P. (2007). Laser-activated rfid-based indoor localization system for mobile robots. pages 4600 –4605.

Zivkovic, Z. and Kröse, B. (2007). Part based people detection using 2D range data and images. In *Int. Conf. on Robotics and Automation (ICRA'07)*, Roma, Italy.

Résumé

Ces travaux de thèse s'inscrivent dans le cadre du projet européen CommRob impliquant des partenaires académiques et industriels. Le but du projet est la conception d'un robot compagnon évoluant en milieu structuré, dynamique et fortement encombré par la présence d'autres agents partageant l'espace (autres robots, humains). Dans ce cadre, notre contribution porte plus spécifiquement sur la perception multimodale des usagers du robot (utilisateur et passants). La perception multimodale porte sur le développement et l'intégration de fonctions perceptuelles pour la détection, l'identification de personnes et l'analyse spatio-temporelle de leurs déplacements afin de communiquer avec le robot. La détection proximale des usagers du robot s'appuie sur une perception multimodale couplant des données hétérogènes issues de différents capteurs. Les humains détectés puis reconnus sont alors suivis dans le flot vidéo délivré par une caméra embarquée afin d'en interpréter leurs déplacements.

Une première contribution réside dans la mise en place de fonctions de détection et d'identification de personnes depuis un robot mobile.

Une deuxième contribution concerne l'analyse spatio-temporelle de ces percepts pour le suivi de l'utilisateur dans un premier temps, de l'ensemble des personnes situées aux alentours du robot dans un deuxième temps. Enfin, dans le sens des exigences de la robotique, la thèse comporte deux volets : un volet formel et algorithmique qui tire pertinence et validation d'un fort volet expérimental et intégratif. Ces développements s'appuient sur notre plateforme Rackham et celle mise en œuvre durant le projet CommRob.